

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 14:34:08 (Search time 1230 Seconds
(Without alignments)
4775.961 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783
1 atccactctgtcttga.....tagagcgtcaagcgaacc 1783

Scoring table:

IDENTITY NTC
Gapop 10.0, Gapext 1.0

Searched: 2190069 seqs, 1647345023 residues

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA: *

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09C_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|---------|----|--------------------|
| C 1 | 88.4 | 5.0 | 3673778 | 12 | US-10-312-841-2 |
| C 2 | 82.4 | 4.6 | 5830 | 12 | US-10-311-455-490 |
| C 3 | 80 | 4.5 | 11996 | 12 | US-10-240-485-45 |
| C 4 | 77.8 | 4.4 | 11996 | 12 | US-10-240-485-45 |
| C 5 | 76.4 | 4.3 | 6641 | 12 | US-10-311-455-288 |
| C 6 | 76.4 | 4.3 | 18154 | 12 | US-10-240-485-287 |
| C 7 | 74.2 | 4.2 | 10329 | 12 | US-10-311-455-206 |
| C 8 | 74.2 | 4.2 | 7597 | 12 | US-10-311-455-206 |
| C 9 | 73.6 | 4.1 | 3673778 | 12 | US-10-312-841-1 |
| C 10 | 72.8 | 4.1 | 6189 | 12 | US-10-240-485-146 |
| C 11 | 72.6 | 4.1 | 5689 | 12 | US-10-240-485-100 |
| C 12 | 72.6 | 4.1 | 5689 | 14 | US-10-239-676-90 |
| C 13 | 72.2 | 4.0 | 6065 | 12 | US-10-311-455-477 |
| C 14 | 71.8 | 4.0 | 5807 | 12 | US-10-311-455-1128 |
| C 15 | 71.8 | 4.0 | 11996 | 12 | US-10-240-485-46 |

| | | | | | | |
|------|------|-----|---------|----|--------------------|--------------------|
| 17 | 71.4 | 4.0 | 5276 | 12 | US-10-311-455-123 | Sequence 123, App |
| C 18 | 71.4 | 4.0 | 19087 | 12 | US-10-311-455-766 | Sequence 766, App |
| C 19 | 70.4 | 3.9 | 3673778 | 12 | US-10-312-841-1 | Sequence 1, App |
| C 20 | 70.4 | 3.9 | 6255 | 12 | US-10-311-455-933 | Sequence 933, App |
| C 21 | 70.4 | 3.9 | 8170 | 12 | US-10-240-453-132 | Sequence 132, App |
| C 22 | 70.2 | 3.9 | 8305 | 12 | US-10-311-455-1542 | Sequence 1542, App |
| C 23 | 69.8 | 3.9 | 40862 | 12 | US-10-311-455-2046 | Sequence 2046, App |
| C 24 | 69.6 | 3.9 | 6071 | 12 | US-10-311-455-298 | Sequence 298, App |
| C 25 | 69.6 | 3.9 | 6270 | 12 | US-10-311-455-1845 | Sequence 1845, App |
| C 26 | 69.4 | 3.9 | 7851 | 12 | US-10-311-455-1734 | Sequence 1734, App |
| C 27 | 69.4 | 3.9 | 3673778 | 12 | US-10-312-841-2 | Sequence 2, App |
| C 28 | 69.2 | 3.9 | 6535 | 12 | US-10-311-455-909 | Sequence 909, App |
| C 29 | 69 | 3.9 | 960 | 14 | US-10-198-846-6381 | Sequence 6381, App |
| C 30 | 69 | 3.9 | 18624 | 12 | US-10-311-455-1676 | Sequence 1676, App |
| C 31 | 68.8 | 3.9 | 14095 | 12 | US-10-311-455-450 | Sequence 450, App |
| C 32 | 68.6 | 3.8 | 17967 | 12 | US-10-311-455-988 | Sequence 988, App |
| C 33 | 68.4 | 3.8 | 18218 | 12 | US-10-311-455-1922 | Sequence 1922, App |
| C 34 | 68.2 | 3.8 | 11996 | 12 | US-10-240-485-45 | Sequence 45, App |
| C 35 | 68 | 3.8 | 5930 | 12 | US-10-311-455-490 | Sequence 490, App |
| C 36 | 68 | 3.8 | 7516 | 12 | US-10-311-455-34 | Sequence 34, App |
| C 37 | 68 | 3.8 | 15548 | 12 | US-10-311-455-2128 | Sequence 2128, App |
| C 38 | 67.8 | 3.8 | 12574 | 12 | US-10-311-455-1290 | Sequence 1290, App |
| C 39 | 67.6 | 3.8 | 5929 | 12 | US-10-311-455-681 | Sequence 681, App |
| C 40 | 67.6 | 3.8 | 6103 | 12 | US-10-311-455-1664 | Sequence 1664, App |
| C 41 | 67.6 | 3.8 | 40862 | 12 | US-10-311-455-2046 | Sequence 2046, App |
| C 42 | 67.4 | 3.8 | 13125 | 12 | US-10-311-455-1199 | Sequence 1199, App |
| C 43 | 67.4 | 3.8 | 13125 | 12 | US-10-240-485-109 | Sequence 109, App |
| C 44 | 67.2 | 3.8 | 5379 | 12 | US-10-311-455-1650 | Sequence 1650, App |
| C 45 | 67.2 | 3.8 | 5379 | 12 | US-10-240-485-130 | Sequence 130, App |

ALIGNMENTS

RESULT 1
US-10-312-841-2/c
Sequence 2, Application US/10312841
Publication No. US20030186277A1
GENERAL INFORMATION:
APPLICANT: Epigenomics AG
TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MK
FILE REFERENCE: E01/1208/WO
CURRENT APPLICATION NUMBER: US/10/312,841
CURRENT FILING DATE: 2002-12-30
NUMBER OF SEQ ID NOS: 2
SEQ ID NO 2
LENGTH: 3673778
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (379615)
US-10-312-841-2

Query Match
Best local similarity 49.4%; Pred. No. 0.00012;
Matches 347; Conservative 0; Mismatches 341; Indels 14; Gaps 4;

| | | | |
|----|---------|---|---------|
| QY | 782 | ATAAGACCACTTACGACACTTTTCTAATAATCTACCAAGGAGGAAATTTTAACTACT | 841 |
| DB | 1400831 | AAAAACCAATTTTAAATGACACTTATACAAATTAACAATTAATAAACAAT | 1400832 |
| QY | 842 | TAGAGAGTATGGAGTTTAAAGCAACATTAAGGGGAGCTTAAATTAATGTGT | 901 |
| DB | 1400831 | AAAAATTTAAATATATATATTTAACTTAACTTTTATTTTAAATATATATC | 1400772 |
| QY | 902 | TGTACACACACACTTCTTGTAGATTTTAAAGAAATTTGATCATCATTAATTT | 961 |
| DB | 1400771 | TAAACCTATATATTTATTAATAATGCAATTAATAAATTAATTCATTAATTT | 1400712 |
| QY | 962 | ATTGCTTATTTAAATTAATTAATGATTAAGATGT-ATCATTAAGATGAGAAACCAATAG | 1020 |

| | | | |
|----|------|---|------|
| Dd | 5089 | TAAAAAAATTTACTCATCAAACTTATCTAAAACCTCCAACTTCCTCAATTATTAATCTAAT | 5038 |
| Oy | 1056 | TTAAGTTTTCTCAGACCCTATATAAACTTGTAAGCGTAAATGTGATCGTAGAAAAA | 1115 |
| Dd | 5029 | TTTATATATTTTCAAACATTTTAAAAATAAACAAAAATGCAAATTAATATAAAAA | 4970 |
| Oy | 1116 | TGTGATGAATTCATNGAAATTATGTAATTTCAANGTCCAAATCCATNGAANT | 1175 |
| Dd | 4969 | ACAAATATATTTACTTTAAAAAAAATCTTAATAAATAAAAAAACCTTAAAAACCAT | 4910 |
| Oy | 1176 | TTAGTCAAAAAGTAACTCAAAAATATTCCTATTTTAAATTTTCAACAATATATAAA | 1235 |
| Dd | 4909 | AACAAAAAT | 4850 |
| Oy | 1236 | TATTCCTTATTTAAATTTTACATATATATTAATTATACCTGGTCACTTTAGATACC | 1295 |
| Dd | 4849 | AACCTCAACTTAAAAATTAATATATATATATATATATATATATATATATATATAT | 4790 |
| Oy | 1296 | ACCAACAATATTAATTAAGTATATATTTTATCTTAAATATTTTGAGATCTCATATAT | 1355 |
| Dd | 4789 | AAAT | 4730 |
| Oy | 1356 | CCTGATATTTATTTATATTTGTGCATATTTTCTATGTTTAGAGTTAACCTTA | 1411 |
| Dd | 4729 | ATTAACATTAT | 4674 |

RESULT 3
US-10-240-485-45
Sequence 45, Application US/10240485
Publication No. US20030148327A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander

```

1 TITLE OF INVENTION: Diagnosis of Diseases Associated with
2 TITLE OF INVENTION: Metastasis
3 FILE REFERENCE: 5013.1007
4 CURRENT APPLICATION NUMBER: US/10/240,485
5 PRIOR FILING DATE: 2002-10-02
6 PRIOR APPLICATION NUMBER: PCT/EP01/03970
7 PRIOR FILING DATE: 2001-04-06
8 PRIOR APPLICATION NUMBER: DE 10019058.8
9 PRIOR FILING DATE: 2000-04-06
10 PRIOR APPLICATION NUMBER: DE 10019173.8
11 PRIOR FILING DATE: 2000-04-07
12 PRIOR APPLICATION NUMBER: DE 10032522.9
13 PRIOR FILING DATE: 2000-06-30
14 PRIOR APPLICATION NUMBER: DE 10043926.1
15 PRIOR FILING DATE: 2000-09-01
16 NUMBER OF SEQ ID NOS: 202
17 SEQ ID NO: 45
18 LENGTH: 11996
19 TYPE: DNA
20 ORGANISM: Artificial Sequence
21 FEATURE:
22 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
23 US-10-240-485-45
24
25 Query Match 4.5% Score 80; DB 12; Length 11996;
26 Best Local Similarity 49.0% Pred. NO. 0.00044;
27 Matches 242; Conservative 0; Mismatches 250; Indels 2; Gaps 1;

```

| | | | |
|----|------|--|------|
| Qy | 912 | ACTACCTTAGAGCTATTTTAAAGAAATGATATCATCACTTAAATATATGGT - CCT | 969 |
| Db | 929 | ATTATATTATATTATATTATATATATATATATATATATATATATATATATATAT | 988 |
| Qy | 970 | TATTTAAATATATGATAAGTGTGATCATTTAAATGAGAAAACCAATAGCCCTCGCT | 1023 |
| Db | 988 | TATATTAATTA | 1048 |
| Qy | 1030 | TGATTTTGAATATATGTTTTCTATGTACTTTCTTCAGCCCTATATATAAAACTTTGTA | 1088 |


```

Db      2741 TTAATATAGAGATGTTTATATTAATATATTTTAAAGAAATAGTAGATAT 2800
QY      995 TCATTAGATGAGAAACCAATAGCTCGTCTGTGATTTTGAATTTATGTTTCTAT 1054
Db      2801 TGAATTAATTTTGGTTAGAGATTTAGATTTATTTTATTTAGATTTTGAAGTTATAT 2860
QY      1055 GTTACTTTTCTCAAGCCTATATATAAACTTTGTA--ATGCTAAATGTAAGCTGAAA 1111
Db      2861 GTTAATTAATGAATTTATTTATTTAGAAAGATATAGAGATTTAAATAGATTAATA 2920
QY      1112 AAAATGTGAATGAATTCATATAGAAATATG--TATTTCAAAGTCCAAATGCATCAAT 1169
Db      2921 GTTATTTTAAATTAATTAATTTTATTTATTTTGTGATTTAAATTAATGATTTAT 2980
QY      1170 AGAATTTAGTACAAAGTACCAAAATATTTCTTATTTTAAATTTTACACAAAT 1229
Db      2981 AATAAGTATAGATTAATTTATTTATTTATTTAGAGATTTATTTATTTAATAATA 3040
QY      1230 TAAAAATATCTCTATTTTAAATTTTACAAATATATATTTATCAGCTGCACCTTAG 1289
Db      3041 GTTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3100
QY      1290 AATACCACCAACATTTAATTTAGTATTTATTTATTTATTTATTTATTTATTTATTT 3149
Db      3101 ATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3160
QY      1350 ATATATCTGATTTATTTATTTA--TATTTGTGATTTATTTCTTATTTATTTAGAGTAA 1407
Db      3161 AATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3220
QY      1408 CTATATCTTGTGCAACCTAGTATTTCAATTTATTTATTTATTTATTTATTTATTT 3280
Db      3221 AATAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3280
QY      1468 TTGAACATTTGTTTAACTTTGTGAGATTTAAAGTAAATTAACATTCAGATTTAG 1527
Db      3281 GTAGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3340
QY      1528 ACCATCTATTTATTTACTCTTGTGCTTT 1557
Db      3341 TATATGAATTTAGGTTTAAATTTATTTT 3370

```

RESULT 6

```

US-10-240-452-36
; Sequence 36, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBERG, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240.452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 36
; LENGTH: 6641
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE: chemically created genomic DNA (Homo sapiens)

```

US-10-240-452-36

Query Match 4.3%; Score 76.4; DB 12; Length 6641;
 Best Local Similarity 45.9%; Pred. No. 0.0015;
 Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;

```

QY      755 AATGTGAGTTACCTTTCTGCACTTATAGAGCACTTACGACACTTTTACATAAT 814
Db      2561 ATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2620
QY      815 ACTCAAGAGAGAGATTTTAACTTATAGAGATTAAGGAGTTAAAGACACAT 874
Db      2621 ATATTAAGTAATTAATTTATTTATTAATTAAGATATATGATTTTAAAGAGTTGATG 2680
QY      875 TAAAGGAGAGTTTAAATTTATGCTGTGTAACACACACTTCTTATAGATTTATTA 934
Db      2661 TGAAGAGTTAGATTTTATTTTATTTTGTATTTTATTAAGATTTATTTATTTATTT 2740
QY      935 GAAATTTGATATCATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 994
Db      2741 TTAATATTAAGATTTTATTTATTAATAATTTTAAAGAAATATTTAGATAT 2800
QY      995 TCATTAGATTTGAGAAACCAATAGCTCGTCTGATTTTGAATTTATGTTTTCAT 1054
Db      2801 TGAATTAATTTTGTGTTTACAGATTTAGATTTATTTTATTTTATTTTATTTATTT 2860
QY      1055 GTTACTTTTCTCAAGCCTATATATAAACTTTGTA--ATGCTAAATGTAAGCTGAAA 1111
Db      2861 GTTATATATGAATTTATTTATTTATTTAGAGATTTATGAGATTTAATAATAGATTAATA 2920
QY      1112 AAAATGTGATATTTCAATTAAGAAATTTAG--TATTTCAAAGTCCAAATGCATCAAT 1169
Db      2921 GTTATTTTAAATTAAGTAAATTTTATTTATTTTGTGATTTAAATTAAGAAATTTAT 2980
QY      1170 AGAATTTAGTACAAAGTACCAAAATATTTCTTATTTTAAATTTTACACAAAT 1229
Db      2981 AATAAGTATAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3040
QY      1230 TAAAAATATCTCTTATTTTAAATTTTACAAATTTATTTATTTATTTATTTATTTATTT 1289
Db      3041 GTTAATTTATTTATTTAAAGATTTATTTAGATTTATTTATTTATTTATTTATTTAT 3100
QY      1290 AATACCACCAACATTTAATTTAGTATTTTATTTATTTATTTATTTATTTATTTATTT 1349
Db      3101 ATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3160
QY      1350 ATATATCTGATTTATTTT--TATTTGTGATTTTCTTATTTTATTTATTTATTTATTT 1407
Db      3161 AATAATTTAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3220
QY      1408 CTATATCTTGTGCAACCTAGTATTTCAATTTATTTATTTATTTATTTATTTATTTATTT 1467
Db      3221 AATAATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3280
QY      1468 TTGAACATTTGTTTAACTTTGTGAGATTTAAAGTAAATTAACATTCAGATTTAG 1527
Db      3281 GTAGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3340
QY      1528 ACCATCTATTTATTTACTCTTGTGCTTT 1557
Db      3341 TATATGAATTTAGGTTTAAATTTATTTT 3370

```

RESULT 7

```

US-10-311-455-227
; Sequence 227, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBERG, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation

```



```
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 227
; LENGTH: 18154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-227
```

```
Query Match 4.3%; Score 76.4; DB 12; Length 18154;
Best Local Similarity 4.2%; Pred. No. 0.0022;
Matches 303; Conservative 0; Mismatches 331; Indels 8; Gaps 2;
```

```
QY 927 TATATAGAGAAATGTATCATCATATTAATTAATTCCTATTTAAATTAAGATA 986
Db 13819 TTTTATTAATTAATTTGATATATTAATTTTAAATTAATTTGCTTT 13878
QY 987 AAGTGTATCATTAAGATTAAGAAACCAATAGTCTGCTGATTTTGAATTATG 1046
Db 13879 TAAATTTATTAATTAAGA-AAATTAATTAATTTGATTAATTAATTAATTT 13937
QY 1047 TTTTCATGTACTCTTCTTCAAGCCATATATAAATTTGTAATGCTAATGCT 1106
Db 13938 TATTTTATTTATTTTATTAATTAATTTGATTAATTAATTTTAAATTAATTT 13997
QY 1107 GGAATAAATGTATGAATGAATCAATGAATTAATGTAATTCAGATCCAAATCCATC 1166
Db 13998 TGATTTGATGTTTGTATTTTAAATTAATTTGATTAATTTTAAATTTTAAATG 14057
QY 1167 AATAGAAATTTAGTACAAAACGTACCAAAAATTTCTCTTATTTAAATTTACACA 1226
Db 14058 TTTTATTTTATTTTGTATTTTATTAATTTTAAATTTTGAATTTTAAAGTTTGTG----- 14112
QY 1227 ATATATAAATATCTCTTATTTTAAATTTACATATAATTAATTAATTCACCTGCACCT 1286
Db 14113 -ATTTGAATATAATTTATTTATTTTAAATTTTAAATTTGATTTTAAATTTTAA 14170
QY 1287 TAGAATACCAACAATATTAATTAATTAATTTTAAATTTTAAATTTTGAATCTC 1346
Db 14171 TAAATTTGAAGTATTTTATTTAGTTTAAATTTTAAATTTTAAATTTTAAATTA 14230
QY 1347 TCAATATATCTGATTTATTTATTTATTTGTCATATTTCTTATGTTTGAAGTAAAC 1406
Db 14231 TTTTAAATTTTATTTTATTTTATTTTATTTTAAATTTTAAATTTTAAATTTTAA 14290
QY 1407 CCTTATATCTGTGCAACTAGTAATTAATTAATTAATTTGTAAGAGACATTCACAT 1466
Db 14291 ATTTTAAAGTTTGTATATAATTAATTTATTAATTTAGATATGCTGTATTTAATG 14350
QY 1467 CTGGAACAATGCTTTTAACTGTTGGAATGTTAAAGTAATTAATTAATTAATTAAT 1526
Db 14351 TTTTATTTTATTTTATTTTATTTTAAATTTGTAATTAATTAATTTTAAATTAATTA 14410
QY 1527 GACCATCTATTAATATATCTCTGTTGCTTTTAAATTAATTAATTTTAAATTTT 1568
Db 14411 TATATATAGGAATATATATGTTTATTTTATTTGTAATTAATTTTAAATTTTAA 14452
```

```
RESULT 8
US-10-311-455-2096/c
; Sequence 2096, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEX, Alexander
```

```
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2096
; LENGTH: 10329
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2096
```

```
Query Match 4.2%; Score 74.2; DB 12; Length 10329;
Best Local Similarity 4.78%; Pred. No. 0.0042;
Matches 313; Conservative 0; Mismatches 333; Indels 9; Gaps 3;
```

```
QY 712 TTAACATAAACAACACCTTCATTCGAGAGTATGAAGAAGAAATGCGAGTACCTT 771
Db 5982 TTCATTTTAAACATTAATTAATTCAGACATCTTATTAATTAATTTATTAATTAAC 5923
QY 772 TCTGAGTTCATAGAGCAACTTACAGACACTTTTACTAAA---ATACATCAAAAGAGA 827
Db 5922 TATTCATTCATTAATTTATATATATATTCATTAATTAATTAATTAATTAATTAATTA 5863
QY 828 ACATTTTACAACTTAGAGAGTATGGAGTTAAAGCAACATTAATTAAGGGAGAGT 887
Db 5862 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5803
QY 888 TAAATTTAATGTTGTAGACACACATCTTATTAATTAATTAATTAATTAATTAATTAATC 947
Db 5802 TTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5743
QY 948 ATCAATTAATTAATTTGCTCTTATTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1007
Db 5742 ACCTTATTAACACAAAAATCTAAATTTTAAATTAATTAATTAATTAATTAATTAATTA 5683
QY 1008 GAAAAACAATAGTCTGCTGCTGATTTTGAATTAATTTGCTATGTTTCTTCTTC 1067
Db 5682 TATAATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 5623
QY 1068 AAGCTTATTAATTAATTTGTAAT-GCTAAATTTGATGCTGAAAAAATGTTGTAATGA 1126
Db 5622 AAACATTAATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTA 5563
QY 1127 TTCATTAATTAATTTGATTTCAAGCCAAATTCATATTAATTAATTAATTAATTAATTA 1186
Db 5562 TTAATAAATTAATTTATTTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5503
QY 1187 CGTAACGAAAAATATTTCTTATTTTAAATTTTAAATTAATTAATTAATTAATTAATTA 1246
Db 5502 CAATACGAAATTAATTTCCCTTAAATAAATA---AAACAACTTAAATAATTAATTAATTA 5447
QY 1247 TTTTAAATTTTAAATTAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 1306
Db 5446 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5387
QY 1307 TAACTATTAATTTATTTCTTATTAATTAATTTTGAATCTGATTAATTAATTAATTA 1361
Db 5386 ATATATTTTCCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5332
```

```
RESULT 9
US-10-311-455-986
```



```

RESULT 11
US-10-240-485-146
; Sequence 146: Application US/10240485
; Publication No. US20030148327A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEBENROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with
FILE REFERENCE: 5013.1007
CURRENT APPLICATION NUMBER: US/10/240,485
PRIORITY FILING DATE: 2002-10-02
PRIORITY FILING DATE: PCT/EP01/03970
PRIORITY FILING DATE: 2001-04-06
PRIORITY APPLICATION NUMBER: DE 10019058.8
PRIORITY FILING DATE: 2000-04-06
PRIORITY APPLICATION NUMBER: DE 10019173.8
PRIORITY FILING DATE: 2000-04-07
PRIORITY APPLICATION NUMBER: DE 10032529.7
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: DE 10043826.1
PRIORITY FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 202
SEQ ID NO 146
LENGTH: 6189
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-146

Query Match
Best Local Similarity 4.1%; Score 72.8; DB 12; Length 6189;
Matches 488; Conservative 0; Mismatches 612; Indels 18; Gaps 3

341 TGGAAATCGGCACAAAGTTCGAACTGTATATTTCTTAACAGTAGAAAAATTTCTA 400
2028 TGAATAGAGTAAATTAAGTTTGAAGATTAATTTTTTTTGAAGTCGTGGTGTTTA 208
401 AGTGTTCAGATTTTGACTTTTCACCAACCAACTGACTTTTGACTTTCTTAATAACA 460
2088 GAATTTGCAAGATGAGAGATTTTATTAATAGATATAGGATTTTTTATATTAATAA 214
461 AACTCATCTTCTACACATGCTGTATGTAAGAAAGTGATCTTGGAAATTTGATGATGCA 520
2148 AATATGAAATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 2207
521 AAGTCAAGTTTGACTTTTCAGTGTGCAATTGACCAATTTTGCTCTGTGGCAATTCGAA 580
2208 TTTATTTGATATTAATTAATTAATTTGATTTGATTTGATTTGATTAATAAGATGTTGGGAA 2266
581 CCTAATTCATGATCAGTGCAGCGCAAACTTGATGTCATGGAAGATCTTAAGAAAAATTC 640
2268 AAGTGTTTAAGACGGAATGTTTTTTTTTTTGTGGTTAAGTATGTTTTTTTGA 2327
641 TTGAAGACTGAGAGAAAAATTTTGTAGTACACAACAAGAAATCTGTTTTTCATAGTCG 700

```

Db 2328 TGAGGTATTGTAGATATGTAAAGAAAGAGGGGAAATATTTTGGATATTGATAGTAT 2387

Qy 701 GACTAGACACATTTAAACATAAAAACACACTTATCGAAGAGATGTGAAGAGAAATGT 760

Db 2388 GAGTGTATTATTTATTTATATGAAAAAATGGAATAGAGCTATTTTAAAGAGAAAGTTT 2447

Qy 761 GCAGTTACCTTCTTCCG-----GTCATAGAGCAACTTACACACCTTTACTAAAA 813

Db 2448 GTGGAAATGTAAAGTATATTTATGATATATGATTTATTTATTTAAATGAATTAATA 2507

Qy 814 TACTACAAAGAGGAAGATTTTAACTTACAGAGTAATGGAGTTTAAAGACACACA 873

Db 2508 TTTTATTATATATTTATTTATGTTTGTATTTAGAGATATATTTGGTGTGAGAAATTTAAG 2567

Qy 874 TTAAAGGGAGTGTAAAAATTATGCTGTGTACACACACTTATAGTATTTATA 933

Db 2568 TTTTATTGAAAGATTAAGTAAATTAAGTATTAATTTATAGATTTTAAAGTTTGA 2627

Qy 934 AGAAATTTGTATCATCATTTATATTTATGCTCTTATTTAAATTTATGATTAAGTTGT 993

Db 2628 AATTAAATTTATTTAAATATATTTGTATATGATTTTGTGATTAAGTTGTGTAT 2687

Qy 994 ATCATTAAGATTGAGAAAACCAATATAGTCCGCTGTGATTTTGAATTTATTTGTTCTA 1053

Db 2688 ATATAGGGGTGTGAGAAAAAGTTGAGGAAAAATTTAGTGTATTAAGTTATTTTGTG 2747

Qy 1054 TGTACTTTTCTTCAAGCTTATTAATACTTTGTAATGCTAAATTTGTATGCTGAAAAA 1113

Db 2748 TAAATGTTTTTATTAAGAAAGATTAATTAATTAATTAATTTATTCGTTATTTTAAAA 2807

Qy 1114 AATGATATGAATTCATATAGAAATTTATGATTTCAAGTCCAAATTCATCATAGAA 1173

Db 2808 AATGATTTAGGTTATATGTAATAAAAAAAGTTATTAAT-----ATTATTTGSA 2858

Qy 1174 ATTTAGTACAAACGTAATCTCAAAAAATTTCTCTTATTTTAAATTTTACACAATATPAA 1233

Db 2859 ATTGAATTAAGTTGAAAAAGTAAAGTAAATAATTTATTTATTTGTAAGAAATCGTAGATAAT 2918

Qy 1234 AATATTCCTATTTTAAATTTTACATATATATAATTATCACTGTCACTTTAGATA 1293

Db 2919 TTTAAATTTATTTGTAAGAAAGTAAAAAATAATTTATTTATGATATATTTGTATTT 2978

Qy 1294 CCACCAACAATATTAATCTAGATTTTATTTCTTATATATTTTGTGATCTCCAAAT 1353

Db 2979 TAAATGATATTTTATTTAGGTATATTTATTTATTTGTTTGGAGAAAGAAAAATTTTGATA 3038

Qy 1354 ATCTGATATTTATTTTATTTATTTGTGCAATTTTCTATGTTTGTAGATTAACCTTATA 1413

Db 3039 GAGTTATTTTATTTATTTATTTAGGTATATTTT--GGATGATATATTTTATTTTATTT 3096

Qy 1414 TCTGTCAACACTGATATCAATATATATGATTTGTA 1451

Db 3097 GTTTGTGCTTTATTTATTTATTTTGTTTAGTTGTA 3134

RESULT 12

US-10-453-100

Sequence 100, Application US/10240453

Publication No. US20030148326A1

GENERAL INFORMATION:

APPLICANT: OLEK, Alexander

APPLICANT: PISENBROCK, Christian

APPLICANT: BERLIN, Kurt

TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA

TITLE OF INVENTION: Transcription

TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated

FILE REFERENCE: 5013.1009

CURRENT APPLICATION NUMBER: US/10/240.453

CURRENT FILING DATE: 2002-10-02

PRIOR APPLICATION NUMBER: PCT/EP01/03973

PRIOR FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 350
SEQ ID NO: 100
LENGTH: 5689
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-240-453-100

Query Match 4.1%; Score 72.6; DB 12; Length 5689;
Best Local Similarity 46.7%; Pred. No. 0.0064;
Matches 249; Conservative 0; Mismatches 282; Indels 2; Gaps 1;

QY 891 AATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 950
DB 1556 AAATATTTTATTTTATTTTATTAATATGTTATTTTATTTTATTTTATTTAT 1615
QY 951 ACATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1008
DB 1616 TTTGATGATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1675
QY 1009 AAAACCAAAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1068
DB 1676 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1735
QY 1069 AGCCTATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1128
DB 1736 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1795
QY 1129 CAATAGAAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1188
DB 1796 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1855
QY 1189 TAACCAAAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1248
DB 1856 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1915
QY 1249 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1308
DB 1916 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1975
QY 1309 AATCTAGATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1368
DB 1976 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 2035
QY 1369 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1421
DB 2036 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 2088

RESULT 13
US-10-239-676-90
Sequence 90, Application US/10239676
Publication No. US2003082609A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIERENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
FILE REFERENCE: 5013.1003
CURRENT FILING DATE: 2002-09-24

PRIOR APPLICATION NUMBER: PCT/EP01/03968
DE 10019058.8
DE 10019173.8
DE 10032529.7
DE 10043826.1
PRIOR FILING DATE: 2001-04-06
PRIOR FILING DATE: 2000-04-06
PRIOR FILING DATE: 2000-06-30
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 228
SEQ ID NO: 90
LENGTH: 5689
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: (1818, 1930..1931, 1934..1935, 1947..1948, 1966)
US-10-239-676-90

Query Match 4.1%; Score 72.6; DB 14; Length 5689;
Best Local Similarity 46.7%; Pred. No. 0.0064;
Matches 249; Conservative 0; Mismatches 282; Indels 2; Gaps 1;

QY 891 AATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 950
DB 1556 AAATATTTTATTTTATTTTATTAATATGTTATTTTATTTTATTTTATTTAT 1615
QY 951 ACATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1008
DB 1616 TTTGATGATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1675
QY 1009 AAAACCAAAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1068
DB 1676 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1735
QY 1069 AGCCTATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1128
DB 1736 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1795
QY 1129 CAATAGAAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1188
DB 1796 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1855
QY 1189 TAACCAAAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1248
DB 1856 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1915
QY 1249 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1308
DB 1916 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1975
QY 1309 AATCTAGATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1368
DB 1976 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 2035
QY 1369 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 1421
DB 2036 TAAATTAATTAATGCTGTGAACCACTACCTTGTAGTATTAAGAAATGTGAATCATC 2088

RESULT 14
US-10-311-455-477/c
Sequence 477, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIERENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
TITLE OF INVENTION: Cytosine methylation

FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 477
LENGTH: 6065
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-477

Query Match 4.0%; Score 72.2; DB 12; Length 6065;
Best Local Similarity 48.1%; Pred. No. 0.0077;
Matches 271; Conservative 0; Mismatches 283; Indels 9; Gaps 2;

QY 811 AAATCTACAAAGAGAGATTTTAACTAGAGAGATGAGAGTTAAAGAGCAAC 870
DB 1247 AAAAAAAAAAATAAATACTTTAAATACAAACAAATCAATTAATTAATTTTCA 1186
QY 871 ACATTAAGGGGAGTGTAAATTAATGTGTGTAAACCACTACTTACTTACTT 930
DB 1187 ATAAATACTTAACCTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 1136
QY 931 ATAAAGAAATTTGATCATCATCATTAATTAATTAATTAATTAATTAATTAAT 990
DB 1135 ATACTTAACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1076
QY 991 TGTATCATTAAGTGAAGAAACCAATAGTCTGCTGCTGCTGCTGCTGCTGCT 1050
DB 1075 ATTTTAAATATATTAATAAACAATAAATTTTCAAACTTTTCAAAATTA 1017
QY 1051 CTATGTACTTTCTTCAAGCTATATTAATAAATTGTAATGTAATGTAATGTA 1110
DB 1016 AATTTATATTTCAAAAAAATAAATCCAAAAAATAAATTAATTAATTAATTA 957
QY 1111 AAAAAATGTAATGAATTAATGAATTAATGAATTAATGAATTAATGAATTA 1170
DB 956 TAACACTTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 897
QY 1171 GAATTTAGTAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTA 1230
DB 896 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 837
QY 1231 AAAAAATTTCTTTATTTTAAATTTTAACTAATTAATTAATTAATTAATTA 1290
DB 836 CCTCTATTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 777
QY 1291 AATACCAACGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1350
DB 776 ATTTCTTTAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 717
QY 1351 TATATCTGATTTATTTATTTAT 1373
DB 716 ACTAAATAATCAATTAATTAATCT 694

RESULT 15
US-10-311-455-1128/c
Sequence 1128, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEX Alexander
APPLICANT: BERLIN, Kurt
APPLICANT: PIERREBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
cytosine methylation

FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1128
LENGTH: 5807
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1128

Query Match 4.0%; Score 71.8; DB 12; Length 5807;
Best Local Similarity 47.4%; Pred. No. 0.0089;
Matches 313; Conservative 0; Mismatches 342; Indels 6; Gaps 3;

QY 818 ACAAAGAGAGATTTTAACTAGAGAGATGAGAGTTAAAGCAACATTA 877
DB 5384 ATAACTAAATAATTAATAATTAATTAATTAATTAATTAATTAATTAAT 5325
QY 878 GGGGAGTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 937
DB 5324 TAAACCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5265
QY 938 AATGTAATCATCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 997
DB 5264 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5208
QY 998 TTAAGATGAGAGAAACCAATAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1057
DB 5207 TACAAAGTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5148
QY 1058 ACTTTTCTCAAGCTATTAATTAATTAATTAATTAATTAATTAATTAATTA 1117
DB 5147 AATTTACCAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5088
QY 1118 TGTATGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1175
DB 5087 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5028
QY 1176 TTAGTACAAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTA 1235
DB 5027 TATTAATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4968
QY 1236 TATCTCTATTTTAAATTTTAACTAATTAATTAATTAATTAATTAATTA 1295
DB 4967 TAAAAAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4908
QY 1296 ACCAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1355
DB 4907 TCGATTTATTTTAACTAATTAATTAATTAATTAATTAATTAATTAATTA 4868
QY 1356 CTGATTTATTTTAAATTTTGTGCAATTTTCTTAATTTTGAAGTTAAACCTTAATATC 1415
DB 4847 ATCAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4789
QY 1416 TTGCTCAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1475
DB 4788 CAATTCACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4729
QY 1476 T 1476
DB 4728 T 4728

Search completed: November 25, 2003, 16:56:43
Job time : 1246 secs

Wed Nov 26 09:11:06 2003

us-09-831-083-1_1.rmpb

Page 10

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 ; Search time 96 Seconds
(Without alignments)
8197.772 Million cell updates/sec

Title: US-09-831-083-1
Perfect score: 1783
Sequence: 1 acccaacttcgactcttga.....tagagcgcacgaactgaacc 1783

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Issued_Patents.NA:*
2: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5C_COMB.seq:*
5: /cgn2_6/prodata/2/ina/5D_COMB.seq:*
6: /cgn2_6/prodata/2/ina/5E_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| C 1 | 73.8 | 4.1 | 20674 | 4 | US-09-641-638-651 |
| C 2 | 70 | 3.9 | 19124 | 2 | US-08-487-8268-13 |
| C 3 | 68.2 | 3.8 | 20674 | 4 | US-09-641-638-651 |
| C 4 | 67.6 | 3.8 | 665 | 2 | US-08-883-795A-36 |
| C 5 | 67 | 3.8 | 3095 | 6 | 5231168-1 |
| C 6 | 66.6 | 3.7 | 665 | 2 | US-08-883-795A-36 |
| C 7 | 66.2 | 3.7 | 6243 | 2 | US-09-056-075-1 |
| C 8 | 62.6 | 3.5 | 19124 | 2 | US-08-487-8268-13 |
| C 9 | 61.6 | 3.5 | 668 | 3 | US-08-998-416-535 |
| C 10 | 61.6 | 3.5 | 6124 | 4 | US-08-213-4198-3 |
| C 11 | 61 | 3.4 | 168575 | 4 | US-09-426-290-1 |
| C 12 | 60.8 | 3.4 | 666 | 3 | US-08-998-416-117 |
| C 13 | 60 | 3.4 | 53332 | 4 | US-09-801-861-3 |
| C 14 | 59 | 3.3 | 2614 | 4 | US-09-004-056-1 |
| C 15 | 58.8 | 3.3 | 827 | 4 | US-08-998-416-535 |
| C 16 | 58.8 | 3.3 | 58407 | 4 | US-08-916-421B-2 |
| C 17 | 57.8 | 3.2 | 10640 | 4 | US-09-417-485D-5 |
| C 18 | 57.8 | 3.2 | 53332 | 4 | US-09-801-861-3 |
| C 19 | 57.2 | 3.2 | 3701 | 4 | US-08-845-258-10 |
| C 20 | 57.2 | 3.2 | 3701 | 4 | US-08-990-571-10 |
| C 21 | 57.2 | 3.2 | 3701 | 4 | US-08-723-142A-10 |
| C 22 | 57.2 | 3.2 | 3701 | 4 | US-09-528-784A-10 |
| C 23 | 57.2 | 3.2 | 3701 | 4 | US-09-569-098A-10 |
| C 24 | 57 | 3.2 | 615 | 3 | US-08-998-416-186 |
| C 25 | 57 | 3.2 | 731 | 1 | US-08-451-405A-2 |
| C 26 | 57 | 3.2 | 2334 | 1 | US-08-062-632-4 |
| C 27 | 56.8 | 3.2 | 615 | 3 | US-08-998-416-186 |

| | | | | | | |
|------|------|-----|--------|---|--------------------|--------------------|
| 28 | 56.8 | 3.2 | 636 | 3 | US-08-998-416-1137 | Sequence 1137, App |
| 29 | 56.8 | 3.2 | 837 | 3 | US-08-998-416-288 | Sequence 288, App |
| 30 | 56.6 | 3.2 | 834 | 3 | US-08-998-416-305 | Sequence 305, App |
| C 31 | 56.6 | 3.2 | 837 | 3 | US-08-998-416-288 | Sequence 288, App |
| C 32 | 56.6 | 3.2 | 854 | 3 | US-08-998-416-534 | Sequence 534, App |
| C 33 | 55.8 | 3.1 | 834 | 3 | US-08-998-416-305 | Sequence 305, App |
| C 34 | 55.4 | 3.1 | 827 | 3 | US-08-998-416-535 | Sequence 535, App |
| C 35 | 55.4 | 3.1 | 1431 | 3 | US-09-316-083-2 | Sequence 2, App1 |
| C 36 | 55.4 | 3.1 | 1431 | 4 | US-09-933-700-2 | Sequence 2, App1 |
| C 37 | 55.2 | 3.1 | 6265 | 4 | US-09-129-112-3 | Sequence 3, App1 |
| C 38 | 54.8 | 3.1 | 860 | 3 | US-08-998-416-287 | Sequence 287, App |
| C 39 | 54.8 | 3.1 | 168575 | 4 | US-09-426-290-1 | Sequence 1, App1 |
| 40 | 54.6 | 3.1 | 1511 | 1 | US-07-991-867B-8 | Sequence 8, App1 |
| 41 | 54.6 | 3.1 | 1511 | 1 | US-08-107-755A-8 | Sequence 8, App1 |
| 42 | 54.6 | 3.1 | 1511 | 2 | US-08-544-332-8 | Sequence 8, App1 |
| 43 | 54.6 | 3.1 | 1511 | 4 | US-09-370-861A-8 | Sequence 8, App1 |
| C 44 | 54.2 | 3.0 | 731 | 1 | US-08-451-405A-2 | Sequence 2, App1 |
| C 45 | 54.2 | 3.0 | 4818 | 3 | US-08-817-926-27 | Sequence 27, App1 |

ALIGNMENTS

```

RESULT 1
US-09-641-638-651/c
; Sequence 651, Appl 1
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BALESTIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET, 051CPI
; CURRENT FILING DATE: US/09/641, 638
; PRIOR APPLICATION NUMBER: 2000-08-16
; PRIOR FILING DATE: US 09/502, 330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133, 200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275, 267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119, 917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURES:
; NAME/KEY: misc feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5' regulatory region
; NAME/KEY: exon
; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6

```

NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1579
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTAA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele

LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5903
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492

OTHER INFORMATION: 10-507-321 : polymorphic base A or C
 NAME/KEY: allele
 LOCATION: 13524
 OTHER INFORMATION: 10-507-353 : polymorphic base C or T
 NAME/KEY: allele
 LOCATION: 13535

Query Match 4.1%; Score 73.8; DB 4; Length 20674;
 Best Local Similarity 51.3%; Pred. No. 8.8e-06;
 Matches 254; Conservative 0; Mismatches 227; Indels 14; Gaps 3;

QY 1084 TTGTGAATGCTAAATGTATGATCTGGAAGAAAATGTGAAGATCAATGAAATATG 1143
 Db 11558 TTTTAAATATTAATATTTTCTTCTAGCTATTAATATTAATTAATTAATTAAT 11499
 QY 1144 TATTCAAGTCCAAAATCCATCAATGAATTTAGTACAAAACGTAACTCAAAATATT 1203
 Db 11498 TAAATATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11439
 QY 1204 CTCTATTTTAAATTTTACACAAATATAA--AATATCTCTATTTTAAATTTTACAT 1261
 Db 11438 AAATTTAATTTAATTTGAAGCAATTAATTAATTAATTAATTAATTAATTAATTA 11379
 QY 1262 AATATAATTTATACCTGTGACCTTTAGAAATACCAACAATATTAATCTTAAATATT 1321
 Db 11378 ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11319
 QY 1322 TTTTCTTAATTAATTTGAAGATCTCTCAATATCTATTTTATTTTATTTGTCGA 1381
 Db 11318 TAAATTAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 11259
 QY 1382 TATTTCTTATGTTTGTAGCTTAACCTTATATCTTGCAACTGATTAATCAATATAT 1441
 Db 11258 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAAATTT 11199
 QY 1442 GAGTTTGTGAAGACACATTTGACATCTTGAACATGCTTTTAACTT-----GTG 1493
 Db 11198 AAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTAATTA 11139
 QY 1494 GAATGTAAAGTAATTAACATTTCAAGATTAAGA-----CCATCTTAAATTAATCTTCT 1549
 Db 11138 AAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTTCC 11079
 QY 1550 TTGTCTTTTAAAAA 1564
 Db 11078 TTAGCTATTAATAA 11064

RESULT 2

US-08-487-826B-13/c
 Sequence 13, Application US/08487826B
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487, 826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29, 655
 REFERENCE/DOCKET NUMBER: NIH121.001CP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 13:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19124 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 US-08-487-826B-13

Query Match 3.9%; Score 70; DB 2; Length 19124;
 Best Local Similarity 44.3%; Pred. No. 4.7e-05;
 Matches 286; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

QY 770 TTTCGAGTTCATTAAGACCACTTACAGACACTTTTCTAATAATCTACAAAGAGAG 829
 Db 4861 TTTCTCTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4802
 QY 830 ATTTTAACAATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAGAGATTAAG 889
 Db 4801 AATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4742
 QY 890 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 949
 Db 4741 AATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4682
 QY 950 CACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1009
 Db 4681 ACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4622
 QY 1010 AAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1069
 Db 4621 ATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4562
 QY 1070 GCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1129
 Db 4561 CAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4502
 QY 1130 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1189
 Db 4501 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4442
 QY 1190 AACTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1249
 Db 4441 AACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4382
 QY 1250 AATTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1309
 Db 4381 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4322
 QY 1310 TACTAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1369
 Db 4321 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4262
 QY 1370 AATTTTGTGATTAATTTCTTATTTTGAAGTTAATTAATTAATTAATTAATTAATTAATTAAT 1415
 Db 4261 AATATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4216

RESULT 3

US-09-641-638-651
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bouguerelet, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET, 051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6
; NAME/KEY: exon
; LOCATION: 7379..7522
; OTHER INFORMATION: exon 7
; NAME/KEY: exon
; LOCATION: 8645..8854
; OTHER INFORMATION: exon 8
; NAME/KEY: exon
; LOCATION: 12254..12340
; OTHER INFORMATION: exon 9
; NAME/KEY: exon
; LOCATION: 12854..13023
; OTHER INFORMATION: exon 10
; NAME/KEY: exon
; LOCATION: 13308..13429
; OTHER INFORMATION: exon 11
; NAME/KEY: exon
; LOCATION: 16567..16667
; OTHER INFORMATION: exon 12
; NAME/KEY: exon
; LOCATION: 16775..16945
; OTHER INFORMATION: exon 13
; NAME/KEY: exon
; LOCATION: 17063..17554
; OTHER INFORMATION: exon 14

NAME/KEY: misc_feature
LOCATION: 17555..20674
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
OTHER INFORMATION: 10-343-231 : deletion of C
NAME/KEY: allele
LOCATION: 4088
OTHER INFORMATION: 12-206-366 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4109
OTHER INFORMATION: 10-343-278 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 4170
OTHER INFORMATION: 10-343-339 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 5303
OTHER INFORMATION: 10-346-23 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
NAME/KEY: allele

```
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 6534
OTHER INFORMATION: 10-347-271 : polymorphic base A or T
NAME/KEY: allele
LOCATION: 6611
OTHER INFORMATION: 10-347-348 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 7668
OTHER INFORMATION: 10-348-391 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8608
OTHER INFORMATION: 10-349-47 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 8658
OTHER INFORMATION: 10-349-97 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 8703
OTHER INFORMATION: 10-349-142 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 8777
OTHER INFORMATION: 10-349-216 : deletion of CTG
NAME/KEY: allele
LOCATION: 8785
OTHER INFORMATION: 10-349-224 : polymorphic base G or T
NAME/KEY: allele
LOCATION: 8926
OTHER INFORMATION: 10-349-368 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12171
OTHER INFORMATION: 10-350-72 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12429
OTHER INFORMATION: 10-350-332 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13341
OTHER INFORMATION: 10-507-170 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 13492
OTHER INFORMATION: 10-507-321 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 13524
OTHER INFORMATION: 10-507-353 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 13535

Query Match 3.8%; Score 68.2; DB 4; Length 20674;
Best Local Similarity 46.5%; Pred. No. 0.0001;
Matches 257; Conservative 0; Mismatches 293; Indels 3; Gaps 1;
```

```
Db 11156 AAAATTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 11215
Qy 1018 TAGTCCTGCTTGAATTTGAATTAATTTGTTTCATGTTACTTTCTCAAGCCATAT 1077
Db 11216 TTAATTTAAATTAACATTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 11275
Qy 1078 AAAAAGTTGATGCTAAATTAATTAATGCTGCAAAAAATGCTGAATTCATAGAAA 1137
Db 11276 TAATTTAACTTAATTTAAATTTAAATTTAAATTTAAATTTAAATTTAAAT 11335
Qy 1138 TTAATGTAATTTCAAGTCCAAAATCCATCAATAGAAATTTAGTACAAAACGTAATCAA 1197
Db 11336 TTAATTTAAATTTTAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTT 11395
Qy 1198 AATATCTCTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1257
Db 11396 AATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 11455
Qy 1258 CAATATTAATTAATTTTACCTGCTGCTTGAATTAACCAACCAATTAATTAATTTGA 1317
Db 11456 TATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 11515
Qy 1318 TATTTATCTCTTA 1330
Db 11516 TATTAATTTTAA 11528
```

```
RESULT 4
US-08-883-795A-36/c
Sequence 36, Application US/08883795A
Patent No. 5985607
GENERAL INFORMATION:
APPLICANT: Delcove, Genevieve
TITLE OF INVENTION: Aways, Gregor
TITLE OF INVENTION: Recombinant DNA Molecules and Expression
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: BERSKIN & PARR
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/883,795A
FILING DATE: 27-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Michelle
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7841-062
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 665 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 32
US-08-883-795A-36
```

Query Match 3.8%; Score 67.6; DB 2; Length 665;
Best Local Similarity 45.8%; Pred. No. 9.4e-05;
Matches 271; Conservative 0; Mismatches 319; Indels 2; Gaps 1;

1045 TGGTTTCTATGTAAGTCTTTCTTCAAGCTTATATAAAAGCTTGTATGCTAATGTATG 1104
Db TTTTATATTTTTCCTGCTGCTTTTAACTATTTATATAAATGAAATTAATAAATATGTA 537
Qy 1105 CTGGAATAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 1164
Db 536 ATATATATATCTTTAT 477
Qy 1165 TCAATAGAAATTTAGTCAAAAGCTAAGCTAATAATTTCTTATTTTAAATTTTGA 1224
Db 476 ATATATATCTTTAT 417
Qy 1225 CAATATATAATATCTCTTATTTTAAATTTTCAATATATATATATATATATATATATATAT 1284
Db 416 TAAATATGTAATATATAACATTTTATATATATATATATATATATATATATATATATAT 357
Qy 1285 TTTAGAAATACCAACCAAT 1344
Db 356 TAAATATGTAATATATAACATTTTATATATATATATATATATATATATATATATATAT 297
Qy 1345 TCTCATATATCTGAATTTTATTTTATTTTGTGCTATTTCTTATGTTTATAGTGA 1404
Db 296 TAAATATTTTATATATAACATTTTATATATATATATATATATATATATATATATATAT 237
Qy 1405 ACCCTTATCTTGTGCAAACTAGTAATTCATATATATAGTTTGTGAGACACATGAC 1464
Db 236 TAAATATTTTATATATAATTTTATAT - ATAAATATTTTATATATAATTTTATAT 179
Qy 1465 ATCTTGAACATGCTTTTAAACCTGTTGGAATCTTAAAGGTATATAAATTCACAGAT 1524
Db 178 TATATAATTTTATATATAATTTTATATATAATTTTATATATAATTTTATAT 119
Qy 1525 ATGACCATCTATATATATCTTCTTGTCTTTTAAATAAGTGCAATGAAATGCTCA 1584
Db 118 TATATAATTTTATATATAATTTTATATATAATTTTATATATAATTTTATAT 59
Qy 1585 TGGTAAAGCTAGAGTCTGCTGCGCTGTGATTCATTCATTTCCAGAT 1636
Db 58 TATATAATTTTATATATAATTTTATATATAATTTTATATAAATTTTATATCTCAT 7

RESULT 5
5231168-1
; Patent No. 5231168
; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JENSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 1:
; LENGTH: 3095
5231168-1

Query Match 3.8%; Score 67; DB 6; Length 3095;
Best Local Similarity 44.1%; Pred. No. 0.00014;
Matches 280; Conservative 0; Mismatches 355; Indels 0; Gaps 0;
Qy 924 AAGTATTTAAGAAATTTGTAATCATCATTAATTAATTTGCTTATTTAAATATATG 983
Db 2377 AAAAAAAAAAAAAAAAAATTAATAAATTTTCTTATATATATGTAACATAATATAT 2436
Qy 984 ATAAAGTTTATCATTAAGATTGAGAAACCAATATGTCCTGCTGATTTTGAATTA 1043
Db 2437 TATTAATATAATATATATATATATATAAAGAAATCTAGAGATTCTGATATATGAAATTA 2496
Qy 1044 TTGTTTCTATGTAATCTTCTTCAAGCTATATATAAATTTGTAATGCAATTTGTAT 1103

Db 2497 ATTCAATGTATATTTATATAGAAAAATATAAATATAATTTCTTTCTTGTGTAT 2556
Qy 1104 GCTGAAAAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 1163
Db 2557 TAACTATCTAAAAATATAATATAAATAAATAATAGAAACATTTACAGTATATTTTATATCT 2616
Qy 1164 ATCAATAGAAATTTAGTCAAAAGCTAAGCTAATAATTTCTTATTTTAAATTTTACA 1223
Db 2617 GAAAGTATACATGTAATATAAATAATTTTCTGCTTTTAAATATAAATTAAGAAATATAT 2676
Qy 1224 ACAATATATAATATCTCTTATTTTAAATTTTCAATATATATATATATATATATATATATAT 1283
Db 2677 ATTTATGTAATTTATTTTAAAGAAATATATATATATATATATATATATATATATATATAT 2736
Qy 1284 CTTTGAATACCAACCAATATATATATATATATATATTTTATTTTATATATTTGAGAT 1343
Db 2737 AAAAAATTTTATATATATATTTTCTCAATAATATATATAATTTATATATATAATA 2796
Qy 1344 CTTCAATATATCTGAATTTTATTTTATTTTGTGCAATATTTTCTTATGTTTATAGATT 1403
Db 2797 CTGATTAATTTTATATATATATTTTATATATATATATATATATATATATATATATATAT 2856
Qy 1404 AACCTTATATCTTGTGCAAACTAGTAATTCATATATATAGTTTGTGAGACACATGTA 1463
Db 2857 AATAAATAATAGAAACCAATATAATCTATGTTCTGATTTATATTTTATAGTAAATATAATCA 2916
Qy 1464 CATCTGAACATGCTTTTAAACCTGTTGGAATGTTAAAGGTATATAAATTCACAGAT 1523
Db 2917 TATATATTTTATATATTTTAAATATTTTAAACCAATGATATATATAATTTTAAATAG 2976
Qy 1524 TATGACCATCTATTAATTAATCTTCTGCTTTT 1558
Db 2977 AAAAAATGATTTTATATATGTCATAAATTTT 3011

RESULT 6
US-08-883-795A-36
; Sequence 36, Application US/08883795A
; Patent No. 5985607
; GENERAL INFORMATION:
; APPLICANT: Delcive, Genevieve
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BERESKIN & PARR
; STREET: 40 King Street West
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5H 3Y2
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/883,795A
; FILING DATE: 27-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Micheline
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 7841-062
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 665 base pairs
; TYPE: nucleic acid

STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: Rh 32
US-08-883-795A-36

Query Match 3.7%; Score 66.6; DB 2; Length 665;
Best Local Similarity 48.0%; Pred. No. 0.00015;
Matches 282; Conservative 0; Mismatches 299; Indels 6; Gaps 3;

QY 959 ATTATGTCCTATTAAATATGATPAAGCTGTATCATTAAGATTGAGAAACCAAT 1018
DB 14 AATGCTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73
QY 1019 AGTCCTGCTGCTGATTTTGAATTAATTAATTAATTAATTAATTAATTAATTA 1078
DB 74 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 133
QY 1079 A-AACTTGTAATGCTAAATGATGCGAAAAAATGCTAATGAATTCATAGAA 1136
DB 134 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 193
QY 1137 ATTAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1196
DB 194 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 253
QY 1197 AAATATTCCTATTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1256
DB 254 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 310
QY 1257 ACATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1316
DB 311 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 369
QY 1317 ATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1376
DB 370 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 429
QY 1377 TGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1436
DB 430 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 489
QY 1437 TATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1496
DB 490 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 549
QY 1497 TGTAAAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1543
DB 550 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 596

RESULT 7

US-09-056-075-1
Sequence 1, Application US/09056075
Patent No. 5955368
GENERAL INFORMATION:
APPLICANT: Johnson, Eric A.
APPLICANT: Bradshaw, Marite
TITLE OF INVENTION: Expression System for Clostridium
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
FILING DATE:
APPLICATION NUMBER: US/09/056,075
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95238
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6243 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: misc feature
LOCATION: 3720..4013
OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
OTHER INFORMATION: plasmid RP4"
US-09-056-075-1

Query Match 3.7%; Score 66.2; DB 2; Length 6243;
Best Local Similarity 47.1%; Pred. No. 0.00022;
Matches 303; Conservative 0; Mismatches 333; Indels 7; Gaps 3;

QY 917 CTTTGAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 976
DB 2768 CTTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2827
QY 977 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1036
DB 2828 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2887
QY 1037 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1096
DB 2888 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2947
QY 1097 ATGTAATGCTG-GAAAAAATGTAATGAATTAATTAATTAATTAATTAATTA 1154
DB 2948 GTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3007
QY 1155 CCAAAATCCATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1214
DB 3008 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3067
QY 1215 AATTTTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1274
DB 3068 ATTTTTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3127
QY 1275 AACTGTCCTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1334
DB 3128 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3185
QY 1335 TTTGAGATCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1394
DB 3186 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3245
QY 1395 TTTGAGTAACTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1454
DB 3246 TTCCGCTTTA--TTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3302
QY 1455 ACATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1514
DB 3303 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3362

QY 1515 ATTCAGATTATGACCATCTATTATATACCTCTTGTCTT 1557
DB 3363 TTTTATTTATTTATTTATTTATTTATTTATTTTAAAGTTT 3405

RESULT 8
US-08-487-826B-13

/ Sequence 13, Application US/08487826B
/ Patent No. 5993827
/ GENERAL INFORMATION:
/ APPLICANT: Sim, Kim L.
/ APPLICANT: Chitnis, Chetan
/ APPLICANT: Miller, Louis H.
/ APPLICANT: Peterson, David S.
/ APPLICANT: Su, Xin-zhaun
/ APPLICANT: Wellens, Thomas E.
/ TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
/ TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
/ NUMBER OF SEQUENCES: 45
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Knudde Martens Olson & Bear
/ STREET: 620 Newport Center Drive 16th Floor
/ CITY: Newport Beach
/ STATE: California
/ COUNTRY: US
/ ZIP: 92660
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/487,826B
/ FILING DATE: 10-SEP-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Israelson, Ned
/ REGISTRATION NUMBER: 29,655
/ REFERENCE/DOCKET NUMBER: NIH121.001CPI
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (619) 235-8550
/ TELEFAX: (619) 235-0176
/ INFORMATION FOR SEQ ID NO: 13:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 19124 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: CDNA
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ US-08-487-826B-13

Query Match 3.5%; Score 62.6; DB 2; Length 19124;
Best Local Similarity 47.4%; Pred. No. 0.0012;
Matches 226; Conservative 0; Mismatches 244; Indels 7; Gaps 1;
QY 906 ACCACCACTACCTTAGTACTATTATAGAAAATGTAATCATACATTATATATTG 965
DB 6547 ACTACTTTATATATATACCAATACCAAGATATGTAATCATATATATATATA 6606
QY 966 TCCTTATTTAAATTAATGATTAAGTTGATCATTTAAGATGAGAAAACCAATAGTCTC 1025
DB 6607 TGAATATATATAATATGATAGATATATAGATAGAGAAAACCAAGCAATATTGCTCT 6666
QY 1026 GTCTTGATTTTGATTTATGTTTCTTCTATGTTACTT-----TCTTCAAGCCTATATA 1078
DB 6667 TTGTGATCTCTAAT 6726
QY 1079 AAAACTTTGATGCTAAATTTGATGCTGCAAAAAATGCTATGATTCATTCAGAAAT 1138
DB 6727 CATATATATATGTTAATG 6786

QY 1139 TATGTAATTCAGAGCCAAATCCATCATATAGAAATTTAGTACAAACGTAACCAAA 1198
DB 6787 TTGATATTTGGATATTTTCTTCTCATTTATATATTTACTTATATATATATATATAT 6846
QY 1199 ATATTCCTGTAATTTTAAATTTTACACACATATATATATTTCTTATTTTAAATTTAC 1258
DB 6847 AAAT 6906
QY 1259 AAT 1318
DB 6907 GATTAATATTTTGTGTTAGAAATTTTAAATTTATATATATATATATATATATATAT 6966
QY 1319 ATTTATTTCTTAAATATTTTGAGATCTCTCAATATATCTATATTTTATATTT 1375
DB 6967 ATATTTTAT 7023

RESULT 9
US-08-998-416-595

/ Sequence 595, Application US/08998416
/ Patent No. 6239264
/ GENERAL INFORMATION:
/ APPLICANT: Philippsen, Peter
/ APPLICANT: Pohlmann, Rainer
/ APPLICANT: Steiner, Sabine
/ APPLICANT: Mohr, Christine
/ APPLICANT: Wendland, Jurgen
/ APPLICANT: Knechtle, Philipp
/ APPLICANT: Reibschung, Corinne
/ TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII
/ TITLE OF INVENTION: AND USES THEREOF
/ NUMBER OF SEQUENCES: 1152
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: No. 6239264artis Corporation
/ STREET: 3054 Cornwallis Road
/ CITY: Research Triangle Park
/ STATE: No 6239264th Carolina
/ COUNTRY: USA
/ ZIP: 27709
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/998,416
/ FILING DATE: 24-DEC-1997
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: CH 0016/97
/ FILING DATE: 31-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Weig, J. Timothy
/ REGISTRATION NUMBER: 38,241
/ REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 919-541-8587
/ TELEFAX: 919-541-8689
/ INFORMATION FOR SEQ ID NO: 595:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 658 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: PAG1408RP.
/ US-08-998-416-595

Query Match 3.5%; Score 61.6; DB 3; Length 658;
Best Local Similarity 47.7%; Pred. No. 0.0013;
Matches 244; Conservative 0; Mismatches 264; Indels 4; Gaps 2;

QY 904 TAACCACACCTACCTTAGTAGATATAGAGAAATGTGATCATCATTAATTAAT 963
DB 135 TAAGCACCACATATGAATATGAATAGTTTATGATCGTTGATATATACCATATTAAGT 134
QY 964 TGGCTTATTTAAATATGATTAAGTGTATCTTTAAGATGAGAAACCAATATGTC 1023
DB 195 TATCTTATTAATTAATTTATTTATTTGATTTATTTTATTTTATTTTATTTTATTT 251
QY 1024 TCGCTGATTTTGAATATATGTTTCTGATGTTTCTTCAAGCCTATATATTAAT 1083
DB 252 ATATTTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 311
QY 1084 TTGTGAATGCTAAATGTATAGCTGAGAAAATGCTGAATGAATCAATGAAATATAGG 1143
DB 312 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 371
QY 1144 TATTTCAAGTCGCAAAATGCAATGCAATGAAATTTAG-TAGAAAACGTAATCAAAATAT 1202
DB 372 TTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 431
QY 1203 TCTCTATTTTAAATTTTACACAAATATTAATAATTTCTCTATTTTAAATTTTACAATA 1262
DB 432 CTATGATATTAATATTAATTAATGATAGATATATTAATGATATTTATTTATTTATTT 491
QY 1263 ATATTAATTTACACCTGTCACCTTTAGATACCAACCAATATTAATTTATTTATTT 1322
DB 492 ATATAGATGTATTTTGTGACAAATATATATATGCTATTTCTATTTATTTATTTATTT 551
QY 1323 TATTTCTAATATTTTGAATCTCTCAATATCTGATATTTATTTATTTATTTATTTATTT 1382
DB 552 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 611
QY 1383 ATTTTCTATGTTTATGAGTTAAACCTTATAT 1414
DB 612 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 643

RESULT 10

US-06-213-419B-3/c
; Sequence 3, Application US/08213419B
; Patent No. 6333406
; GENERAL INFORMATION:
; APPLICANT: Inselburg, J. et al.
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: J11-002CNCB
; CURRENT APPLICATION NUMBER: US/08/213,419B
; CURRENT FILING DATE: 1994-03-14
; PRIOR APPLICATION NUMBER: US 07/870,506
; PRIOR FILING DATE: 1992-04-17
; NUMBER OF SEQ. ID NOS: 20
; SOFTWARE: Patentln Ver. 2.0
; SEQ. ID NO 3
; LENGTH: 6124
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2407)..(2439)
; NAME/KEY: CDS
; LOCATION: (2598)..(3404)
; NAME/KEY: CDS
; LOCATION: (3580)..(3720)
; NAME/KEY: CDS
; LOCATION: (3850)..(5835)
; US-06-213-419B-3

Query Match 3.5%; Score 61.6; DB 4; Length 6124;
Best Local Similarity 47.7%; Pred. No. 0.0017;
Matches 210; Conservative 0; Mismatches 229; Indels 1; Gaps 1;
QY 958 AATATATGCTTATTAATTAATATGATTAAGTGTATCATTTAGATTGAGAAAACCAA 1017

DB 2579 AATGTTGTGATGCGTAAATAATATATATATTTATATATATATATATTTTACA 2520
QY 1018 TAGTCTGCTCTGATTTTGAATATTTGTTTCTATGTTACTTTTCTTCAAGCTATAT 1077
DB 2519 TATTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2460
QY 1078 AAAAATTTGATGCTAAATGTATGCTGAAAAATGCTATGATGATCAATAGAAA 1137
DB 2459 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2401
QY 1138 TTATGATATTTCAAGTCCAAATCCATCATATAGAAATTTAGTACAAAACGTAATCAAA 1197
DB 2400 ATATGAAATATTAAGTTATATATATATTTATAGAACTTATTTTATTTGAAATATGTTGAT 2341
QY 1198 AATATCTCTATTTTAAATTTTACAAATATTAATAATTTCTCTATTTTAAATTTTA 1257
DB 2340 TATTTAATTTTCTTTATAGTTATATATATGATTTTATATATATATATATATTTATTT 2281
QY 1258 CAATATATATTTATTCACCTGTCACCTTTAGATACCAACCAATATTAATTTACTTAGA 1317
DB 2280 TTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2221
QY 1318 TTTTATTTCTTAATATTTTGAATCTCTCATATATCTGATATTTATTTATTTTCT 1377
DB 2220 TATATATATATTAATTTAAATATATATATATATATTTATTTATTTATTTATTTATTT 2161
QY 1378 GTCATATTTTCTTATGTTT 1397
DB 2160 TTTATTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2141

RESULT 11

US-09-426-290-1/c
; Sequence 1, Application US/09426290
; Patent No. 6410712
; GENERAL INFORMATION:
; APPLICANT: Berglind Ran Olafsdottir
; APPLICANT: Jeffrey Guicher
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
; FILE REFERENCE: 2345.2001-000
; CURRENT APPLICATION NUMBER: US/09/426,290
; CURRENT FILING DATE: 1999-10-25
; NUMBER OF SEQ. ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ. ID NO 1
; LENGTH: 168575
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (21181)..(21403)
; NAME/KEY: CDS
; LOCATION: (95252)..(95430)
; NAME/KEY: CDS
; LOCATION: (101753)..(101996)
; NAME/KEY: CDS
; LOCATION: (110324)..(110439)
; NAME/KEY: CDS
; LOCATION: (124058)..(124278)
; NAME/KEY: CDS
; LOCATION: (127009)..(127130)
; NAME/KEY: CDS
; LOCATION: (128910)..(129139)
; US-09-426-290-1

Query Match 3.4%; Score 61; DB 4; Length 168575;
Best Local Similarity 49.1%; Pred. No. 0.0031;
Matches 250; Conservative 0; Mismatches 250; Indels 9; Gaps 3;
QY 923 TAAGATATTAAGAAATGTGAATCATCATTAATTAATGCTTATTTAAATTAAT 982
DB 109335 TAAATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 109276

| | | | |
|----|--------|---|--------|
| OY | 983 | GATTAAGTGTGTCATTAGATGGAGAAACCAATAAGTCCTGGCTTGATTTTGAAAT | 1042 |
| Db | 109275 | AATATCATTAATATATTTATTAATAATTTATATATTCATATTTAGATATATTCATATA | 109216 |
| OY | 1043 | ATTGTTTTCTATGTACTTTTCTCGAAGCATAATATAAAAACTTGTATGCTAAATGTGA | 1102 |
| Db | 109215 | TTAATATAGACATATTTATTTATATATATAATAATATATATATATATATAGTTATATATA | 109156 |
| OY | 1103 | TGCTGAAAAAAATGCTGATATGATATCCATAGAATTTGSGATTTCCAAAGTCCAAATC | 1162 |
| Db | 109155 | TATAT---ATATATTAATATAATATATATATATATATATATATATATATATATAT--- | 109103 |
| OY | 1163 | CATCAATAGCAATTTTAGTCAAAACGTAACGCMAAACATATTTCTGTATTTTAAATTTTAC | 1222 |
| Db | 109102 | TATATATTAATATTTAGATATADAATATAATATATATATATATATATATATATATATAT | 109043 |
| OY | 1223 | AACAATATATAAAATATTTCTTATATTTTAAATTTTCAATATATATATTTATCACCTGTCA | 1282 |
| Db | 109042 | TATATATATATAATATATAATATACTATATATATATATATATATATATATATATATATAT | 108983 |
| OY | 1283 | CCTTGAATACCAACCAACATATPATCTAGATAT--TTATATCTTAATAATTTTGA | 1340 |
| Db | 108982 | ACATTAATATTTTATATATATATATTTTATATATAACATATATATATATATATATATATAT | 108923 |
| OY | 1341 | GATCTCTCAATATATCTGAATTTATTTTATATATTTTGTCATATTTTCTTATGTTTTAGA | 1400 |
| Db | 108922 | CGGGAATCTGAATTTTGTATGATGTATTTTTTAAATTTTATTCGATAGGTTTTGG | 108863 |
| OY | 1401 | GTTTAACCTTATATCTTGCTCAACATAGT | 1429 |
| Db | 108862 | GGAAACAGTTGGTGTTTTGGTTGTCATGAAT | 108934 |

RESULT 12

US-08-998-416-1137/c
Sequence 1137, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philipssen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Weidland, Jürgen
APPLICANT: Knechtke, Philipp
APPLICANT: Reibschung, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPII
TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587

```

1 :
2 : TELEEXP: 919-5541-8669
3 : INFORMATION FOR SEQ ID NO: 1137
4 : SEQUENCE CHARACTERISTICS:
5 :     LENGTH: 536 base pairs
6 :     TYPE: nucleic acid
7 :     STRANDEDNESS: single
8 :     TOPOLOGY: linear
9 : MOLECULE TYPE: DNA (genomic)
10 : ORIGINAL SOURCE:
11 : ORGANISM: PAg1692ARP
12 :
13 : US-08-998-415-1137

```

| | | | | |
|-----------------------|--------------|-------------------|-----------------|-------------|
| Query Match | 3.4%; | Score 60.8; | DB 3; | Length 636; |
| Best Local Similarity | 47.2%; | Pred. No. 0.0019; | | |
| Matches 232; | Conservative | 0; | Mismatches 247; | Indels 1 |

| | | | |
|----|------|--|------|
| QY | 1047 | TTTTCTGATGTAAGTCTTCTCTCCAGCCGATPAAAAAAGCTTGATGCGTAATGTGATGCT | 1106 |
| Db | 613 | TTCTTTAAATATTTTATTATTTATTTAAATTAAGTAATAATATTATTAATAATTAAT | 554 |
| QY | 1107 | GGAAAAAATGCTGTAATGAATTCATAGAAATTGTGTAITTCAGAGTCCAAAATCCATC | 1166 |
| Db | 553 | TTATTTAAATPAAATPAAATPAAATPAAATGAATATATTTTAAATPAAATPAAATTAAT | 494 |
| QY | 1167 | AATGAATTTAGTACAAAACGTACTCAAAAATATCTCTATTTPAAATTTACACA | 1226 |
| Db | 493 | AATAAGAAATTTAAAGTTAAATTTATTTTATTAATATCTTATPAAAAAGTTAAATAT | 434 |
| QY | 1227 | ATATPAAAAATTTCTCTTAATTTTAAATTTTACAATAATPAAATTTATCAGCTGCACCTT | 1286 |
| Db | 433 | ATPAAATCAACATTAATTTTAAAAAATAGATATATATATAAAAAATPAAATTTACAAAT | 374 |
| QY | 1287 | TAGATTCACCAACCAATPAAATTAATAGTATGATTTTATCTTATPAAATTTTGAGATCTC | 1346 |
| Db | 373 | TTAATTAATTAATTAATCTTTATATATPAAATATATTTATTTTATAAA----- | 326 |
| QY | 1347 | TCAATATATCTGAATATTTATTTATATTTGTGCATATTTCTTATGTTTAAAGGTATAC | 1406 |
| Db | 325 | -CAATTAATAATATATATATTAATTAATGATATACATTTTAAATATTTATTTAAAGAAAT | 267 |
| QY | 1407 | CCTATATATCTTGCGCAACTGTATTCATATPAAATGATTTGTGAAGACACATTCACAT | 1466 |
| Db | 266 | AATAATATCTATATATATTTTAAATACATATTTTAAATTTGACATGACATPAAATAGAT | 207 |
| QY | 1467 | CTTGAAACATTTGGTTTAAACCTTGTTGGAATGTTAAAGTATATAAAACATTCAGATAT | 1526 |
| Db | 206 | TCATATPAAATATATTTTATTTATATATATPAAATATTAATATGATGATTAAGTAAT | 147 |
| QY | 1527 | GACCATCTATTA 1538 | |
| Db | 146 | ATATATATATTA 135 | |

RESULT 13

```

US-09-801-861-3
/ Sequence 3, Application US/09801861
/ Patent No. 6492154
/ GENERAL INFORMATION:
/ APPLICANT: YAN, Chunhua et al.
/ TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
/ TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
/ TITLE OF INVENTION: THEREOP
/ FILE REFERENCE: CLO01098
/ CURRENT APPLICATION NUMBER: US/09/801,861
/ CURRENT FILING DATE: 2001-03-09
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO. 3
/ LENGTH: 53332
/ TYPE: DNA
/ ORGANISM: Human
US-09-801-861-3

```


Query Match 3.4%; Score 60; DB 4; Length 53332;
 Best Local Similarity 47.6%; Pred. No. 0.0043;
 Matches 177; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

QY 1112 AAAATGCTAATGATTCATAGAAATTAAGTATTCAGAAATTCATCAATAG 1171
 DB 31296 AAAATGCAATCAGTATATATTTATATATATTTATATATTTATATAT 31355
 QY 1172 AAATTGTGCAAAACGTAATCTCAAAATATCTCTATTTTAAATTTTCAACATATA 1231
 DB 31356 TTATATATGTAATATATATATATATATATATATATATATATATATAT 31415
 QY 1232 AAAATATCTCTATTTTAAATTTTACAAATATATATATATATATATATAT 1291
 DB 31416 TATATATTTATATATTTTATATATATTTATATATTTTATATATTTT 31475
 QY 1292 TACCAACAACATTTTAACTTATGATATTTTATCTTAAATTTTGAGATCTCTCAAT 1351
 DB 31476 ATATATATTTTATATATATATATATATATATATATATATATATATAT 31535
 QY 1352 ATATCTGATATTTTATATATTTTGTGCAATTTTCTTANGTTTGAAGTTAACTTA 1411
 DB 31536 ATATTTATATATTTTATATATTTTATATATTTTATATATTTTATATAT 31595
 QY 1412 TATCTTGCTCAACTAGTAAATTCATATATATGAGTTTGGAAGACATTCATCTTGA 1471
 DB 31596 TATATTTTATATATTTTATATATTTTATATATTTTATATATTTTATAT 31655
 QY 1472 AACATTTGTTT 1483
 DB 31656 TATTTTTTTTT 31667

RESULT 14
 US-09-004-056-1
 ; Sequence 1, Application US/09004056A
 ; Patent No. 656586
 ; GENERAL INFORMATION:
 ; APPLICANT: Calgene LLC
 ; FILE OF INVENTION: Plant Expansin Promoter Sequences
 ; TITLE REFERENCE: 125
 ; CURRENT APPLICATION NUMBER: US/09/004, 056A
 ; EARLIER FILING DATE: 1998-01-07
 ; EARLIER APPLICATION NUMBER: 60034914
 ; NUMBER OF SEQ ID NOS: 1
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2614
 ; TYPE: DNA
 ; ORGANISM: Gossypium hirsutum
 ; FEATURES:
 ; NAME/KEY: Promoter
 ; LOCATION: (930)
 ; OTHER INFORMATION: unknown nucleotide
 ; FEATURE:
 ; NAME/KEY: Promoter
 ; LOCATION: (947)
 ; OTHER INFORMATION: unknown nucleotide
 ; FEATURE:
 ; NAME/KEY: Promoter
 ; LOCATION: (956)
 ; OTHER INFORMATION: unknown nucleotide
 ; US-09-004-056-1

Query Match 3.3%; Score 59; DB 4; Length 2614;
 Best Local Similarity 47.4%; Pred. No. 0.0048;
 Matches 211; Conservative 0; Mismatches 230; Indels 4; Gaps 1;
 QY 883 AGTGTAAATTAATGTTGTAACACACACACCTTTGTAAGATTTTAAGAAATTTG 942
 DB 133 ATTTTATATTTATTTTATTTTATATATTTTATATTTTATATATTTTATAT 192

QY 943 TAATCATCATTAATTAATTTGCTTATTTTAAATTTGATTAAGTTGATCATTTAG 1002
 DB 193 AATTTCCATTAACCTTAATTTTCTTAAATTAATTTATTCATTAATAAGAAATTA 252
 QY 1003 ATGGAAGAAACCAATATGTC---CTGCTGATTTTGGATTTTGTCTTCTATGTTA 1058
 DB 253 ATGCTAACCTTTTAAATTAATTCAGAAATTCGTTATTCGAAATTTATGTTCTG 312
 QY 1059 CTTTCTTCAAGCTTATATTAATTAATTTGTAATGCTAAATTTGATGCTGGA 1118
 DB 313 ATATATGTAAGAAATTAATTTTAAATTTTATCAATTAATTTTAAATTAATTA 372
 QY 1119 GTAATGATTCATTAAGAAATTAATGATTTTCAAGTCCAAATTCATCATGAAATTTA 1178
 DB 373 TTTCAATTTTCAACATTAATTTTGAAGATTAAGAACTAAATTAATTAATGA 432
 QY 1179 GTACAAAACGTAACTCAAAATATATCTTATTTTAAATTTTACAACAATATA 1238
 DB 433 TACCAAACTAAACCTTAATTTATTAATTAAGTTCAAAATTAATTAATTA 492
 QY 1239 TCTCTTATTTTAAATTTTCAATTAATTAATTTATCATCTGTCACCTTTAGAT 1298
 DB 493 CCAATATCAACCTTAATTTAGATGAGGCCATTTTAAATTAATTAAGAAATGAACT 552
 QY 1299 AACAAATTAATTAATTTATTTT 1323
 DB 553 AAAATATTTTAAAGTAATTTAT 577

RESULT 15
 US-08-998-416-535
 ; Sequence 535, Application US/08998416
 ; Patent No. 6239264
 ; GENERAL INFORMATION:
 ; APPLICANT: Philippsen, Peter
 ; APPLICANT: Pohlmann, Rainer
 ; APPLICANT: Steiner, Sabine
 ; APPLICANT: Mohr, Christine
 ; APPLICANT: Wendland, Jurgen
 ; APPLICANT: Knechtle, Philipp
 ; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
 ; TITLE OF INVENTION: AND USES THEREOF
 ; NUMBER OF SEQUENCES: 1152
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: No. 6239264artis Corporation
 ; STREET: 3054 Cornwallis Road
 ; CITY: Research Triangle Park
 ; STATE: No. 6239264th Carolina
 ; COUNTRY: USA
 ; ZIP: 27709
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/998,416
 ; FILING DATE: 24-DEC-1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: CH 0016/97
 ; FILING DATE: 31-DEC-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: PR/5-30306/A/CGC1976
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 535:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 827 base pairs

```

;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAC373BP
;
; JS-08-998-416-535

```

Query Match 3.3%; Score 58.8; DB 3; Length 827;
Best Local Similarity 49.1%; Pred. No. 0.0047;
Matches 216; Conservative 0; Mismatches 217; Indels 7; Gaps 2;

| | | | |
|----|------|--|------|
| QY | 1040 | ATTATTTGTTTCTAGTACTTTCTCTCAAGCCTATATATAAAACCTTGAATGCTAAAT | 1099 |
| | | | |
| Db | 166 | ATTGTGGCAGCTTAATTTTATTTATTTAATGTATATATCTATTTACATMAAAACCTT | 225 |
| | | | |
| QY | 1100 | GTATGCTGGAAAAAATGCTGAATGAATCAATAGAAATTAATGATTTCCAAAGCCAAA | 1159 |
| | | | |
| Db | 226 | TTAATAATGTTATATAATATAAATAAATTTCTATATGATAT-TTATTTAAATAGATTTTA | 284 |
| | | | |
| QY | 1160 | ATCCATCAATAGAAATTTAGTACAAAACGTAACTCAAAAATTCCTTAATTTAAATTT | 1219 |
| | | | |
| Db | 285 | ATTATTTATTTATTTTAAATATATACATTTTATATAAATATGATTAATTAAGTTTATTTA | 344 |
| | | | |
| QY | 1220 | TGCAACAAATATAAAAATATCTCTATTTTAAATTTTACAATATATATATATACCTG | 1279 |
| | | | |
| Db | 345 | TATTAAGATGATATATTAATTTATTTATATATAATTTATTTATTTACTCATGATATATAT | 404 |
| | | | |
| QY | 1280 | TCACCTTTAGAAATCCACCAACAATATAATTAATCTATGATATTTTATCTTATATAATTTTG | 1339 |
| | | | |
| Db | 405 | AATTAATTTAAATGACCTTTCATATATTTTATTTTATGCTAGAAATATCTATTTTAA | 464 |
| | | | |
| QY | 1340 | AGATCTCCATATATCTGATAT-T-----TATTTATATTTGTGTGATATTTCTATAG | 1393 |
| | | | |
| Db | 465 | TAGCTAACCTTTAATGATGATATTACTACCTACTAAATTTTAACCTAATATATATATTT | 524 |
| | | | |
| QY | 1394 | TTTTAGATTTAACCTTTATATCTGTGTCAACTAGTAATCAATATATAGAGTTGTGAG | 1453 |
| | | | |
| Db | 525 | AGAAATACTTTAAATCTAAATATTTATTTATCTAATAGATATATTAATTTATCTTTTAA | 584 |
| | | | |
| QY | 1454 | GAACATTTGACATCTTGAAA | 1473 |
| | | | |
| Db | 585 | TTATTAATTTAAATCATTTATA | 604 |
| | | | |

Search completed: November 25, 2003, 15:19:52
Job time : 99 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 / Search time 366 Seconds

(without alignments)
13150.541 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaacttcgaccttga.....tagagcgatcaagctgaacc 1783

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq. 19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1 | 1783 | 100.0 | 1909 | 21 | AAA40223 | V. faba sucrose bi |
| 2 | 91.8 | 5.1 | 8056 | 25 | ABZ10246 | Haematopoietic cel |
| 3 | 91.4 | 5.1 | 8056 | 25 | ABZ10246 | Haematopoietic cel |
| 4 | 89 | 5.0 | 8056 | 25 | ABZ10100 | Haematopoietic cel |
| 5 | 84.4 | 4.7 | 8056 | 25 | ABZ10100 | Haematopoietic cel |
| 6 | 82.4 | 4.6 | 5930 | 24 | ABJ32517 | Human immune syste |
| 7 | 80 | 4.5 | 11996 | 24 | ABJ34492 | Human metastasis a |
| 8 | 79.8 | 4.5 | 700 | 22 | AAH93026 | Human inflammatory |

| | | | | | | | |
|---|----|------|-----|-------|----|----------|--------------------|
| C | 9 | 77.8 | 4.4 | 11996 | 24 | ABJ34493 | Human metastasis a |
| | 10 | 76.4 | 4.3 | 6641 | 24 | ABJ54336 | Chemically treated |
| | 11 | 76.4 | 4.3 | 6641 | 24 | ABJ32315 | Human immune syste |
| | 12 | 76.4 | 4.3 | 18154 | 24 | ABJ32254 | Human immune syste |
| | 13 | 76.2 | 4.3 | 19634 | 25 | ABZ10015 | Haematopoietic cel |
| | 14 | 74.2 | 4.2 | 10329 | 24 | ABJ34123 | Human immune syste |
| | 15 | 74 | 4.2 | 7597 | 24 | ABJ33013 | Human immune syste |
| | 16 | 74 | 4.2 | 32392 | 24 | ABJ56203 | AMEPV genome fragm |
| | 17 | 73.8 | 4.1 | 20674 | 21 | AAO58017 | Archidonic acid m |
| | 18 | 73 | 4.1 | 19634 | 25 | ABZ10261 | Haematopoietic cel |
| | 19 | 72.8 | 4.1 | 6189 | 22 | AAJ46800 | Tumour suppressor |
| | 20 | 72.8 | 4.1 | 6189 | 24 | ABJ34593 | Human metastasis a |
| | 21 | 72.6 | 4.1 | 7667 | 22 | AAJ46334 | Tumour suppressor |
| | 22 | 72.6 | 4.1 | 5689 | 22 | AAJ45384 | Chemically pretrea |
| | 23 | 72.6 | 4.1 | 5689 | 22 | AAJ46426 | Tumour suppressor |
| | 24 | 72.6 | 4.1 | 5689 | 24 | ABK28225 | DNA transcripion |
| | 25 | 72.6 | 4.1 | 14919 | 22 | AAJ45606 | Tumour suppressor |
| | 26 | 72.2 | 4.0 | 6065 | 24 | ABJ32804 | Human immune syste |
| | 27 | 72 | 4.0 | 7025 | 24 | ABK40059 | Human chemically p |
| | 28 | 71.8 | 4.0 | 7025 | 24 | AAJ63350 | Chemically pretrea |
| | 29 | 71.8 | 4.0 | 5807 | 24 | ABJ33155 | Human immune syste |
| | 30 | 71.8 | 4.0 | 11996 | 24 | ABJ34493 | Human metastasis a |
| | 31 | 71.4 | 4.0 | 5276 | 24 | ABJ32150 | Human immune syste |
| | 32 | 71.4 | 4.0 | 19087 | 24 | ABJ32793 | Human immune syste |
| | 33 | 71.4 | 4.0 | 19634 | 25 | ABZ10162 | Haematopoietic cel |
| | 34 | 71 | 4.0 | 16258 | 24 | ABJ70376 | Chemically treated |
| | 35 | 71 | 4.0 | 16258 | 24 | ABK40038 | Human chemically p |
| | 36 | 70.4 | 3.9 | 6255 | 24 | ABJ32960 | Human immune syste |
| | 37 | 70.4 | 3.9 | 8170 | 24 | ABK28258 | DNA transcripion |
| | 38 | 70.4 | 3.9 | 14820 | 24 | ABN80146 | Human chemically m |
| | 39 | 70.2 | 3.9 | 8305 | 24 | ABJ33569 | Human immune syste |
| | 40 | 70 | 3.9 | 19124 | 18 | AAJ72882 | Plasmodium var-7 g |
| | 41 | 70 | 3.9 | 19124 | 21 | AAJ98287 | Plasmodium var-7 g |
| | 42 | 69.8 | 3.9 | 3101 | 11 | AAQ02047 | Sequence encoding |
| | 43 | 69.8 | 3.9 | 40862 | 24 | ABJ34073 | Human immune syste |
| | 44 | 69.8 | 3.9 | 50000 | 24 | ABJ56201 | AMEPV genome fragm |
| | 45 | 69.8 | 3.9 | 50000 | 24 | ABJ56202 | AMEPV genome fragm |

ALIGNMENTS

| | |
|----------|--|
| RESULT 1 | |
| AAA40223 | |
| ID | AAA40223 standard; DNA, 1909 BP. |
| AC | AAA40223; |
| XX | |
| DT | 02-NOV-2000 (first entry) |
| XX | |
| DE | V. faba sucrose binding protein DNA fragment. |
| XX | |
| KW | Soybean; sucrose binding protein; SBP; plant seed; transgenic plant; |
| KW | seed-specific expression; ds. |
| XX | |
| OS | Vicia faba. |
| XX | |
| FT | |
| FT | Key |
| FT | CDS |
| FT | Location/Qualifiers |
| FT | 1784..1909 |
| FT | /*tag= a |
| FT | /product= "SBP" |
| FT | /partial |
| XX | |
| PD | WO200026386-A2. |
| XX | |
| PD | 11-MAY-2000. |
| XX | |
| PF | 27-OCT-1999; |
| XX | |
| PR | 99WO-DE03432. |
| XX | |
| PR | 04-NOV-1998; |
| XX | |
| PA | 98DE-1052195. |
| XX | |
| XX | (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR. |

PI Heim U, Weber H;
XX WPI; 2000-365631/31.
DR P-PSDB; AAB10028.
XX

PT Expression cassette for expressing genes in plant seeds, useful for
PT producing enzymes or pharmaceuticals, includes the promoter from a
PT sucrose-binding protein-related gene

XX
PS Disclosure; Fig 1; 24pp; German.

XX
CC This invention describes a novel expression cassette (I) for expressing
CC genes in plant seeds which comprises (i) the promoter of the
CC sucrose-binding protein (SBP)-like seed protein (II), (ii) optionally the
CC sequence for a signal peptide, particularly from (II), (iii) the gene
CC (III) to be expressed and (iv) a 3'-termination sequence. (I) and a
CC plasmid containing the expression cassette (IV) are used for the
CC expression of homologous or heterologous genes in the seeds of
CC transformed plants, particularly genes that alter the storage properties
CC and germination capacity of the seeds. Alternatively, transgenic plants
CC that express altered or new products in their seeds are selected, grown
CC to establish stable lines and the resulting products (e.g. enzymes,
CC pharmaceuticals or proteins that contain essential amino acids) are
CC extracted. Also insertion of (III) in the antisense orientation may be
CC used to reduce or switch off expression of particular genes. (I) provide
CC seed-specific expression (in cotyledons and endosperm), with stable
CC expression at high level. They provide an overall increase in expression
CC rate, improve utilization of the developmental period of the seed and
CC can overcome the effect of co-suppression. This sequence encodes a
CC fragment of the Vicia faba (soybean) sucrose binding protein (SBP) which
CC is described in the method of the invention.

XX
SQ Sequence 1909 BP; 638 A; 317 C; 281 G; 673 T; 0 other;

Query Match 100.0%; Score 1783; DB 21; Length 1909;
Best Local Similarity 100.0%; Pred. No. 3.5e-296;
Matches 1783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

| | | | |
|----|-----|---|-----|
| QY | 1 | ATCCAACTTCGATCTTGAATCTCTGCTGTCACAACTGTTGGAAGAGCTTGAAGCT | 60 |
| DB | 1 | ATCCAACTTCGATCTTGAATCTCTGCTGTCACAACTGTTGGAAGAGCTTGAAGCT | 60 |
| QY | 61 | TTTGAAGAAAGCTTTGATACAGCTTTGAGACTTTCTTGAATTAAGCTTGGAACTGA | 120 |
| DB | 61 | TTTGAAGAAAGCTTTGATACAGCTTTGAGACTTTCTTGAATTAAGCTTGGAACTGA | 120 |
| QY | 121 | TTGAACCTACGTCGAAAGCTGCTCCAGAGTTCTAACCAAAATTCGCTTGGAGGCCA | 180 |
| DB | 121 | TTGAACCTACGTCGAAAGCTGCTCCAGAGTTCTAACCAAAATTCGCTTGGAGGCCA | 180 |
| QY | 181 | AAATTTATGAGTACCTTCAAGTTTCATGAGAGCTGCTTCAAAAGATTATTAAGTAATCC | 240 |
| DB | 181 | AAATTTATGAGTACCTTCAAGTTTCATGAGAGCTGCTTCAAAAGATTATTAAGTAATCC | 240 |
| QY | 241 | CATATTTTAAAGAGAGTTCTGCTCCGCAATGCTTAGATCTCATTTGAATCTACACT | 300 |
| DB | 241 | CATATTTTAAAGAGAGTTCTGCTCCGCAATGCTTAGATCTCATTTGAATCTACACT | 300 |
| QY | 301 | CTTGTGTGAGAGTTCTTCCAGAAATCACTTGATCAATGCTGAAGTTTGAAGTTTGA | 360 |
| DB | 301 | CTTGTGTGAGAGTTCTTCCAGAAATCACTTGATCAATGCTGAAGTTTGAAGTTTGA | 360 |
| QY | 361 | CTGAACCTGATATTTCTTAACAGTTTGAAGAAATTTCTAAGTGTGAAATTTTGA | 420 |
| DB | 361 | CTGAACCTGATATTTCTTAACAGTTTGAAGAAATTTCTAAGTGTGAAATTTTGA | 420 |
| QY | 421 | TTCCAAAGCAAACTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 480 |
| DB | 421 | TTCCAAAGCAAACTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 480 |
| QY | 481 | CTTGATGAAATGATTTCTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 540 |
| DB | 481 | CTTGATGAAATGATTTCTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 540 |

| | | | |
|----|------|--|------|
| QY | 541 | AGTNGCAATGACCAATTTTCTCTTGTGCAATTTCCAAACCTAATTTGATGATCAGTG | 600 |
| DB | 541 | AGTNGCAATGACCAATTTTCTCTTGTGCAATTTCCAAACCTAATTTGATGATCAGTG | 600 |
| QY | 601 | CTGCAAACTTGAATGATGAGAAAGTCTTATGAGAAATTTCTTGAAGCTGAGAGGAAA | 660 |
| DB | 601 | CTGCAAACTTGAATGATGAGAAAGTCTTATGAGAAATTTCTTGAAGCTGAGAGGAAA | 660 |
| QY | 661 | TTTGTGTGATCAACAAAGAAATCTGTTTCTTCAATGCGAGATGACATTAACATAA | 720 |
| DB | 661 | TTTGTGTGATCAACAAAGAAATCTGTTTCTTCAATGCGAGATGACATTAACATAA | 720 |
| QY | 721 | AACACCACTTCAATGAGAGTGAATGAGAGAAATGTCAGTTCCTTTCGAGCTT | 780 |
| DB | 721 | AACACCACTTCAATGAGAGTGAATGAGAGAAATGTCAGTTCCTTTCGAGCTT | 780 |
| QY | 781 | CATAGAGCACTTACAGACCTTTTCTAATAATACACAAAGAGAGATTTTAAACAAC | 840 |
| DB | 781 | CATAGAGCACTTACAGACCTTTTCTAATAATACACAAAGAGAGATTTTAAACAAC | 840 |
| QY | 841 | TTAGAGAGTATGAGAGTTAAGAGCAACATTAAGGGGAGTGTAAATTAATGTG | 900 |
| DB | 841 | TTAGAGAGTATGAGAGTTAAGAGCAACATTAAGGGGAGTGTAAATTAATGTG | 900 |
| QY | 901 | TTGTAACCACTTACCTTTAGTATGATTAAGAAATTTGTAATCATCATTAATAT | 960 |
| DB | 901 | TTGTAACCACTTACCTTTAGTATGATTAAGAAATTTGTAATCATCATTAATAT | 960 |
| QY | 961 | TATGTCTTATTTAAATTTGATGATTAAGTGTATCATTAAGATTGAGAAAACCAATAG | 1020 |
| DB | 961 | TATGTCTTATTTAAATTTGATGATTAAGTGTATCATTAAGATTGAGAAAACCAATAG | 1020 |
| QY | 1021 | TCTGCTGTTGATTTTGAATTTATTTTCTTATGATTTCTTCAAGCTTATTAATA | 1080 |
| DB | 1021 | TCTGCTGTTGATTTTGAATTTATTTTCTTATGATTTCTTCAAGCTTATTAATA | 1080 |
| QY | 1081 | AACCTTGAATGCTAAATTTGATGCTGAAATAATGTAATGATTAATTAAGAAATTA | 1140 |
| DB | 1081 | AACCTTGAATGCTAAATTTGATGCTGAAATAATGTAATGATTAATTAAGAAATTA | 1140 |
| QY | 1141 | TGGATTTTCAAGTCCAAATTCATCAATAGAAATTTGTAACAAAGCTAATCAAAAT | 1200 |
| DB | 1141 | TGGATTTTCAAGTCCAAATTCATCAATAGAAATTTGTAACAAAGCTAATCAAAAT | 1200 |
| QY | 1201 | ATTCTCTTATTTAAATTTTCAACAAATTAATAATTTCTTATTTTAAATTTTCAAA | 1260 |
| DB | 1201 | ATTCTCTTATTTAAATTTTCAACAAATTAATAATTTCTTATTTTAAATTTTCAAA | 1260 |
| QY | 1261 | TAAATTAATTTATCAGCTGTCACCTTTAGAAATCCACCAAGATTTAATCTTAGATAT | 1320 |
| DB | 1261 | TAAATTAATTTATCAGCTGTCACCTTTAGAAATCCACCAAGATTTAATCTTAGATAT | 1320 |
| QY | 1321 | TTTATTTCTTAATTTTGAATTTCTGCAATATATATTTTATTTATTTATTTATTTAT | 1380 |
| DB | 1321 | TTTATTTCTTAATTTTGAATTTCTGCAATATATATTTTATTTATTTATTTATTTAT | 1380 |
| QY | 1381 | ATATTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 1440 |
| DB | 1381 | ATATTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 1440 |
| QY | 1441 | TGAGTTTGAAGAGCACTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 1500 |
| DB | 1441 | TGAGTTTGAAGAGCACTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA | 1500 |
| QY | 1501 | AAAGGTAATTAACCAATGAGATTTGACATCTATTAATTAATCTCTTGTCTTTTAA | 1560 |
| DB | 1501 | AAAGGTAATTAACCAATGAGATTTGACATCTATTAATTAATCTCTTGTCTTTTAA | 1560 |
| QY | 1561 | AAAGGTAATTAACCAATGAGATTTGACATCTATTAATTAATCTCTTGTCTTTTAA | 1620 |
| DB | 1561 | AAAGGTAATTAACCAATGAGATTTGACATCTATTAATTAATCTCTTGTCTTTTAA | 1620 |

QY 1621 AATTCATTTCCAGATGCTAGAACTGCGCACTACGAATATTAGTCAATAGACACGATG 1680
 DB 1621 AATTCATTTCCAGATGCTAGAACTGCGCACTACGAATATTAGTCAATAGACACGATG 1680
 QY 1681 TTAACACAGCTCCCTGTCATGTTTTTGGCATATATTCGCTCTCTTTCTTTCTTAC 1740
 DB 1681 TTAACACAGCTCCCTGTCATGTTTTTGGCATATATTCGCTCTCTTTCTTTCTTAC 1740
 QY 1741 GTATTAACCAATGAATATTATTAAGCGCATCAAGCTGAAC 1783
 DB 1741 GTATTAACCAATGAATATTATTAAGCGCATCAAGCTGAAC 1783

RESULT 2
 ID AB210246/C
 ID AB210246 standard; DNA; 8056 BP.
 XX
 AC AB210246;
 XX
 DT 16-JAN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX
 KM Human; haematopoietic cell proliferation disorder; cytostatic;
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KM cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200277272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002MO-EP03401.
 XX
 PR 26-MAR-2001; 2001US-278333P.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guefig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Jesche R, Leu E;
 PI Lewin A, Lipscher E, Mader S, Model F, Meller V, Otto T;
 PI Pelet C, Schwobe I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 XX
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated Cpg
 PT dinucleotides -
 XX
 PS Claim 28; SEQ ID 386; 1179p; English.
 XX
 CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated Cpg
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiating between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables

CC a highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.
 XX

Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;
 Query Match 5.1%; Score 91.8; DB 25; Length 8056;
 Best Local Similarity 45.0%; Pred. No. 6.7e-07;
 Matches 481; Conservative 0; Mismatches 572; Indels 16; Gaps 3;

QY 707 ACACATTAACATTAACACACACCTTCATTGAGAGGATGAGAGCAATAGCGACTT 766
 DB 2648 ATATTTTAATAAATAAATATACATTTATTTAATAACAAATTTAATAACAAATTCATTT 2789
 QY 767 ACCCTTCGAGCTTCATTAAGACCACTTACAGACACTTTACTAAATCTACAAAGAG 826
 DB 2788 AATTTATTAATAAATAAATAACATTAATAATTTTTTTTATTTTATTAATATTTATTTAC 2729
 QY 827 AAGATTTTACACACTTAGAAGATTAAGGAGTTAAAGACACACATTAGCGGAGTG 886
 DB 2728 AATATATTTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2669
 QY 887 TTAATAATTAATGTTGTTGACACACACACCTTGTAGTATTAATAAATAATGTAT 946
 DB 2668 TTTAAATTTTAAATTTTAC-----ATTTTTTTTATTTTTTTTTTAAATTTATTTAT 2614
 QY 947 CATCACTTAATTAATTTGTCCTTATTTAAATTAAGATTAAGATTAAGATTTG 1006
 DB 2613 ATTTATTAATTTTTCATTTATTTTATTTTATTTTAAATTTTATTTAATTTTAA 2554
 QY 1007 AAAAAACCAATAGCCCTGCTGCTTATTTTGAATATTTGTTCTATGTTCTTTCTT 1066
 DB 2553 ATTTTATTTATTTATTTAATTAATTTATTTTATTTTATTTTATTTTATTTTATTTAT 2494
 QY 1067 CAAGCCTATTAATAAATCTTGTATGCTAAATTTGATGCTGGAATAAATGTAATGA 1126
 DB 2493 TATTTTAAACAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2434
 QY 1127 TTCAATAGCAATTAATGCTATTTTCAAGTCGCAATTCATCAATAGCAATAA 1186
 DB 2433 AATTAATAAATAATTTATTTTATTTTATTTTCAAAAAATAAATAAATAAATAAATAA 2374
 QY 1187 CGTAACGAAAAATGCTCTTATTTTAAATTTTACACAAATTAATAAATAATTCCTAT 1246
 DB 2373 ATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2314
 QY 1247 TTTAAATTTTACAAATTAATAAATTTATGACCGTGCACCTTTAGAAATACCAACAATAT 1306
 DB 2313 AATATATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2254
 QY 1307 TAAATCTAGATTTTATTTCTTAATAATTTTGAGATCTGCATATATCTGATTTAT 1366
 DB 2253 TAAAAAATAAATTTTATTAATAAATTTTATTTTATTTTATTTTATTTTATTTTATTT 2194
 QY 1367 TTTATATTTTGTGCAATTTTCTTATGTTTATAGATTAACCTTATATCTGTGCAACT 1426
 DB 2193 TTTAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2134
 QY 1427 AGTAAATTA-----ATTAATGAGTTTGTGAAGACACATTTGACATCTTGAACATT 1477
 DB 2133 AAAAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2074
 QY 1478 GGTTTTAACTGTTGGAATGTAAAGGATTAATAACCTGAGATTAATGACCAATAT 1537
 DB 2073 TTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2014
 QY 1538 AATTAATCTGCTTGTCTTTTAAATAAGTGCAATGAATGCTCTATGTTAGCTAGAG 1597
 DB 2013 TAAATTTTATTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1954
 QY 1598 TGTCTGTGCGCGGTGATATCAATTCATTCAGATGTAAGAACTGCACTAGAGAA 1657
 DB 1953 TTTTATTTATTTATTTATTAATTTTATTTATTTTATTTTATTTTATTTTATTTATTA 1894

QY 1658 TAATTAGTCATAGACACGCTATGTAAACAGCCTCCCTGCATGTTTTCATATAT 1717
 DB 1893 AAAAAAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1834
 QY 1718 TCGGCTCTTCCTT-CTCTCAGCTTAAACATATATAT 1764
 DB 1833 AACAAATTTTATTTATTTATTTAAATTAATTAATTAATTAATTAAT 1785

RESULT 3
 AB210246
 ID AB210246 standard; DNA; 8056 BP.
 XX
 AC AB210246;
 DT 16-JUN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.
 XX
 KM Human; haematopoietic cell proliferation disorder; cytostatic;
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KM cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W020027272-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 26-MAR-2002; 2002WO-EP03401.
 XX
 PR 26-MAR-2001; 2001US-278333P.
 XX
 PA (EPIS-) EPIGENOMICS AG.
 XX
 PI Berlin K, Braun A, Distler J, Guertig D, Howe A, Mueller J;
 PI Olek A, Piepenbrock C, Agorjan P, Grabs G, Lesche R, Leu F;
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
 PI Peter C, Schwope I, Ziebarth H;
 XX
 DR WPI; 2003-018942/01.
 PT Detecting and differentiating between hematopoietic cell proliferative
 PT disorders, comprises contacting a target nucleic acid with a reagent
 PT that distinguishes between methylated and non-methylated CpG
 PT dinucleotides -
 XX
 PS Claim 28; SEQ ID 386; 117bp; English.
 XX
 CC The present invention describes a method for detecting and
 CC differentiating between haematopoietic cell proliferative disorders
 CC associated with at least 1 gene and/or their regulatory regions in a
 CC subject. The method comprises contacting a target nucleic acid in a
 CC biological sample obtained from the subject with at least 1 reagent,
 CC which distinguishes between methylated and non-methylated CpG
 CC dinucleotides within the target nucleic acid. AB209861 to AB211118
 CC represent specifically claimed nucleotide sequences from the present
 CC invention. Oligonucleotides from the present invention can be used: for
 CC differentiating between healthy haematopoietic cells and proliferative
 CC disorder haematopoietic cells; for differentiating between acute
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
 CC determining the cytosine methylation state and/or single nucleotide
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
 CC related sequences and their complements; and as primers for the
 CC amplification of haematopoietic cell proliferation disorder related
 CC DNA sequences. The nucleotide sequences from the present invention can
 CC also be used for detecting a predisposition to, differentiation between
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
 CC haematopoietic cell proliferative disorders. The present method enables
 CC a highly specific classification of haematopoietic cell proliferative
 CC disorders allowing for improved and informed treatment of patients.
 XX
 SQ Sequence, 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;

Query Match 5.1%; Score 91.4; DB 25; Length 8056;
 Best Local Similarity 45.7%; Pred. No. 7.9e-07;
 Matches 355; Conservative 0; Mismatches 421; Indels 1; Gaps 1;

QY 799 ACACCTTTACTAAATACACAGAGAGATTTTACACACTTACAGAGATATGAGAG 858
 DB 626 AAAAAATTTTAAAAATATTTTAAATTAATTAATTAATTAATTAATTAATTA 685
 QY 859 TTAAGAGCAACACATTAAGGGAGTGTAAATTAATGATGTGACACACCACTACT 918
 DB 686 ATATTAATAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 745
 QY 919 TTAGTATATTAATTAAGAAATTTGATTCATCATTAATTAATTAATTAATTA 978
 DB 746 TTTTAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 805
 QY 979 TTATG-ATAAGTGTATCATTAAGATTAAGTGAACCAATTAAGTCCGCTGATTTT 1037
 DB 806 AATTTTAATAATGATTTTAAATTAATTAATTAATTAATTAATTAATTTT 865
 QY 1038 GAATTAATTTGTTTCTATGTTACTTTCTTCAAGCCTATATAAACTTGTATGCTAA 1097
 DB 866 TTAATGATATAAAATTTGTAAATTTTAATTAATTAATTAATTAATTTGTTGTTT 925
 QY 1098 TTGATAGCTGAAAAAATGATGATGATGATGATGATGATGATGATGATGATG 1157
 DB 926 GTAATTAATTTTAAATTAATTAATTTTATTTTATTAATTAATTAATTAATTA 985
 QY 1158 AATTCATCAATTAAGAAATTTAGTACAAAGCTACTCAAAATTTCTTATTTAAAT 1217
 DB 986 TTAATTAATTTTAAATTAATTAATTAATTAATTAATTTTATTAATTTTAAATTTGAAA 1045
 QY 1218 TTACAAACAATTAATAAATATTCCTATTTTAAATTTTAAATTAATTAATTAATTAATCACC 1277
 DB 1046 TTTATTAATGTAATAAATTAATTAATTTGTTTATTAATTTGTTTATTTTAAATGTTTAA 1105
 QY 1278 TGTCACTTTAAGATTAACCAACATATTAATTAATTAATTAATTTTATTTATTAATTT 1337
 DB 1106 TGTATTAATTAATTTGTTTAAATTTTAAATTTTAAATTTGTAATTTGAAATTAATTAATTT 1165
 QY 1338 TGAGATCTGCAATTAATGATGATTTTATTTATTTGTCATTAATTTCTTATGTTT 1397
 DB 1166 TTTATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTA 1225
 QY 1398 AGAGTTAACCTTAATTAATTTGTTCAACCTTAATTAATTAATTAATTAATTTGTAAGAGACA 1457
 DB 1226 TTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTG 1285
 QY 1458 CATGACATCTTGAACATTAATGTTTAACTTTGTAATGTTAAAGTAATTAATTAATTT 1517
 DB 1286 AATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTAATTAATTA 1345
 QY 1518 CAGATTAATGACATCTTAATTAATTAATTTGCTTTTAAATTAATTAATTTGTAAGAGACA 1574
 DB 1346 ATTTAAATTTTAAATTTTAAATTAATTAATTTTAAATTAATTAATTTGTAAGAGACA 1402

RESULT 4
 AB210100/C
 ID AB210100 standard; DNA; 8056 BP.
 XX
 AC AB210100;
 DT 16-JUN-2003 (first entry)
 XX
 DE Haematopoietic cell proliferation disorder related DNA sequence #240.
 XX
 KM Human; haematopoietic cell proliferation disorder; cytostatic;
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
 KM cytosine methylation state; gene; ds.
 XX
 OS Homo sapiens.

```

XX      WO20027272-A2.
PN      03-OCT-2002.
XX      26-MAR-2002; 2002WO-EP03401.
FD      26-MAR-2001; 2001US-278333P.
XX      (EPIC-) EPIGENOMICS AG.
XX      Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI      Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI      Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI      Pelet C, Schwöpe I, Ziebarth H;
XX      WPI; 2003-018942/01.
XX      Detecting and differentiating between hematopoietic cell proliferative
PT      disorders, comprises contacting a target nucleic acid with a reagent
PT      that distinguishes between methylated and non-methylated CpG
PT      dinucleotides -
XX      Claim 28; SEQ ID 240; 117bp; English.
XX      The present invention describes a method for detecting and
CC      differentiating between hematopoietic cell proliferative disorders
CC      associated with at least 1 gene and/or their regulatory regions in a
CC      subject. The method comprises contacting a target nucleic acid in a
CC      biological sample obtained from the subject with at least 1 reagent,
CC      which distinguishes between methylated and non-methylated CpG
CC      dinucleotides within the target nucleic acid. AB209861 to AB21118
CC      represent specifically claimed nucleotide sequences from the present
CC      invention. Oligonucleotides from the present invention can be used: for
CC      differentiating between healthy hematopoietic cells and proliferative
CC      disorder hematopoietic cells; for differentiating between acute
CC      lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC      determining the cytosine methylation state and/or single nucleotide
CC      polymorphisms (SNPs) of hematopoietic cell proliferation disorder
CC      related sequences and their complements; and as primers for the
CC      amplification of hematopoietic cell proliferation disorder related
CC      DNA sequences. The nucleotide sequences from the present invention can
CC      also be used for detecting a predisposition to, differentiation between
CC      subclases, diagnosis, prognosis, treatment and/or monitoring of
CC      hematopoietic cell proliferative disorders. The present method enables
CC      a highly specific classification of hematopoietic cell proliferative
CC      disorders allowing for improved and informed treatment of patients.
XX      SQ      Sequence 8056 BP; 371 A; 371 C; 371 G; 3603 T; 0 other;
Query Match      5.0%; Score 89; DB 25; Length 8056;
Best Local Similarity 48.2%; Pred. No. 2e-06;
Matches 309; Conservative 0; Mismatches 330; Indels 2; Gaps 2;
QY      929 TTTAAGAAATGTGATCATCATTAATATGTCCTATTAAATATGATTA 988
DB      2543 TTTTATTAATTAATTAATTTTATTTTATTTTTCGTAATATTT-TAATATTTTAA 2485
QY      989 GTTGATTCATTAGATTGAGAAAACCAATAGTCCGTGATTGATTTGAAATTAATGTT 1048
DB      2484 ACGAAATTAATAAAAAATATATAATAATAAAAAATTAATTTTAAATAATAATA 2425
QY      1049 TTCTATGCTACTTTCTTCAAGCCTATATAAAAACCTTTGTAATGCTAAATGTAAGCTGG 1108
DB      2424 AATTATTTTATTTTATTTTATTTTGAATAATAATAATTAATAATTAATTTTAAATA 2365
QY      1109 AAAAAATGTGTAATGATTCATAGAAATTAATGATTTTCAAGCCAAATTCATCA 1168
DB      2364 AAAAAATTAATTAATAATAATAATAATTTTAAATACCAAAATTAATTAATTTT 2305
QY      1169 TTAGAATTTAGTACAAACGTAACTCAAAATTTCTCTTAATTTTAAATTTTACACA-A 1227
DB      2304 TAAAAATAATAATAATAATAATAATTTTAAATTTTAAATTTTAAATTTTAAATAATA 2245

```

```

QY      1228 TATAAAAATATCTCTTAATTTTAAATTTTACAAATATATATTAACCTGCACCTTT 1287
DB      2244 AATTTTATTAATAATTAATTTTTTTTTTTTTTTTTTTTATTTTTCGTTTTTTAAATTT 2185
QY      1288 AGAATACCCACCAATATTAATTAATCTAGATATTTTATTTCTTAATTAATTTGAGATCT 1347
DB      2184 TTAATTAATAATTAATTTTTTTTTTTTTTTTTTTTTTATTTTAAATAATAATAATAATTA 2125
QY      1348 CAATATATCTGATTAATTTATTTATTTATTTGTCATATTTTCTTATGTTTAAAGTTAAAC 1407
DB      2124 TAAAAATTAATTAATAATTTTAAATTTATTTATTTTAAATTTTAAATTTTAAATAA 2065
QY      1408 CTTAATATCTTGCTCAACTAGTAATTCATATATAGTTTGTGAAGACACATTTGACATC 1467
DB      2064 ATTTTATTAATAATAATAATAATTTTAAATAATATATTTTAAATATATTTTAAATTTT 2005
QY      1468 TTGAACATTTGTTTAACTTGTGAAATGTTAAAGTAATAAAACATTCGAATTAAG 1527
DB      2004 ATTAATTTTAAATAATAATAATAATTAATTAATTTTATTTTAAATAATTAATTTT 1945
QY      1528 ACCATCTAATTAATATACCTTCCTTCCTTTTAAAAAGTGT 1568
DB      1944 TTTATATTAATAATTAATTAATTTTTTTTTTTTAAATAATTTT 1904
RESULT 5
AB210100
ID      AB210100 standard; DNA; 8056 BP.
XX      AB210100;
AC      16-7AN-2003 (first entry)
XX      DT      16-7AN-2003 (first entry)
XX      DE      Hematopoietic cell proliferation disorder related DNA sequence #240.
XX      KW      Human; hematopoietic cell proliferation disorder; cytostatic;
XX      KW      gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
XX      KW      cytosine methylation state; gene; ds.
OS      Homo sapiens.
XX      PN      WO20027272-A2.
XX      PD      03-OCT-2002.
XX      PF      26-MAR-2002; 2002WO-EP03401.
XX      PR      26-MAR-2001; 2001US-278333P.
XX      PA      (EPIC-) EPIGENOMICS AG.
XX      Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI      Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI      Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI      Pelet C, Schwöpe I, Ziebarth H;
XX      WPI; 2003-018942/01.
XX      Detecting and differentiating between hematopoietic cell proliferative
PT      disorders, comprises contacting a target nucleic acid with a reagent
PT      that distinguishes between methylated and non-methylated CpG
PT      dinucleotides -
XX      Claim 28; SEQ ID 240; 117bp; English.
XX      The present invention describes a method for detecting and
CC      differentiating between hematopoietic cell proliferative disorders
CC      associated with at least 1 gene and/or their regulatory regions in a
CC      subject. The method comprises contacting a target nucleic acid in a
CC      biological sample obtained from the subject with at least 1 reagent,
CC      which distinguishes between methylated and non-methylated CpG
CC      dinucleotides within the target nucleic acid. AB209861 to AB21118

```

CC represent specifically claimed nucleotide sequences from the present
CC invention. Oligonucleotides from the present invention can be used: for
CC differentiating between healthy haematopoietic cells and proliferative
CC disorder haematopoietic cells; for differentiating between acute
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC determining the cytosine methylation state and/or single nucleotide
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder
CC related sequences and their complements; and as primers for the
CC amplification of haematopoietic cell proliferation disorder related
CC DNA sequences. The nucleotide sequences from the present invention can
CC also be used for detecting a predisposition to, differentiating between
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of
CC haematopoietic cell proliferative disorders. The present method enables
CC a highly specific classification of haematopoietic cell proliferative
CC disorders allowing for improved and informed treatment of patients.

Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 other;

| | | | | |
|-----------------------|--------------|--------------------|-----------------|--------------|
| Query Match | 4.7%; | Score 84.4; | DB 25; | Length 8056; |
| Best Local Similarity | 46.7%; | Pred. No. 1.2e-05; | | |
| Matches 369; | Conservative | 0; | Mismatches 416; | Indels 5; |
| | | | | Gaps 3 |

| | | | |
|----|------|---|------|
| QY | 779 | TTTCATGAGGCAACTTACGACACCTTTACCTAAATCTCCAAAGGGAAGATTTTACA | 838 |
| Db | 274 | TTTAAATAAAAAATTTTATTTTATTTTAAATATTTTAAAAAATATTAATTA | 333 |
| QY | 839 | ACTTACAGACAGTAATGGAGTT-AAAGGCAACATTAAAGGGAGCTTAATAAT | 897 |
| Db | 334 | ATPAAATAAAAATTAATTAATTTAAAAAATAAAAAATGAAATATTAATTTTA | 393 |
| QY | 898 | GGTGCTGAACCACTACCTTGTAGTATTTAAGAAATTTGATCATCATAT | 957 |
| Db | 394 | AAATTAATTAATTAATTAATTTTAATTAATTTTATTTTAAAAAATATTA | 453 |
| QY | 958 | AATTAATTCCTTAATTTAAATATGATTAAGTTGTTCATTAAATGAGAAACCAA | 1017 |
| Db | 454 | AAATTAATTTTAAAAATTAATTAATTTTAAAAAATTTAAAAATTAATATTT-A | 512 |
| QY | 1018 | TAGTCCTCGCTGATTTTGGATTAATGTTTCTATGTACTTTCTTCAAGCCATAT | 1077 |
| Db | 513 | TAAATATTAATTTTAAATTAATTAATTAATTTTAATTAATTTTAAGCTTTAT | 572 |
| QY | 1078 | AAAACTTGTATGTCATAATTTGATGCTGAAAAAATGTGTATGAATTCATTA | 1137 |
| Db | 573 | AAATTAATTAATTAATTAATTAATTAATTTAAAAAATTTAAATTTTAAAAATTT | 632 |
| QY | 1138 | TATAGTATTCAAAGTCCAAATCCATCATTAATTAATTAAGTACAAACGTAC | 1197 |
| Db | 633 | TTTAAAAATTAATTTTAAATTAATTTATTAATTTATTAATTAATTAATTAATTA | 692 |
| QY | 1198 | AATATTCCTTAATTTAAATTTTACAACAATTAATTAATTTCTATTTTAAATTTTA | 1257 |
| Db | 693 | AAAAATTTTATTTTAAAAATTTAAATTAATTAATTAATTTATTAATTAATTAATTTT | 752 |
| QY | 1258 | CAATAATTAATTTTACCGTGCACCTTGAGTATACCAACAAATTAATTAATCTG | 1317 |
| Db | 753 | ATTTTAAATTTAAATTAATTAATTAATTAATTAATTTATTAATTTAAATTAATTTA | 812 |
| QY | 1318 | TATTTATTCCTTAATTAATTTTGAATCTCCAAATTAATTAATTAATTTTAAATTTGT | 1377 |
| Db | 813 | TAAAGCATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTTTATTTTAAACGT | 872 |
| QY | 1378 | GTCAATTTTCTTANGTTTGAATTAACCTTA--TACTGTGCACAAGTAATTC | 1437 |
| Db | 873 | ATPAAATTTCTTAATTTTAAATTTTAATTTTAATTTTATTTTTCGTGTTTGTGATTA | 932 |
| QY | 1435 | AAATATAGAGTTTGGAAGACACATTCACATCTGAAACNTGTTTAAACCTGTGG | 1494 |
| Db | 933 | ATTTTATTTTAAAAATTAATTTTATTTTATTTTAAAGCAATTAATTAATTAAT | 992 |
| QY | 1495 | AATGTTAAAGTAATTAATTAATTCAGAAATTAAGCACTCATTAATTAATCTCTTGTGC | 1554 |
| Db | 993 | TTTTTAAAAATTAATTTTAAAAATTTTATTAATTTTAAATTTTGAATTAATTT | 1053 |

| | | | |
|----|------|-----------|------|
| QY | 1555 | TTTTAAAAA | 1564 |
| | | | |
| Db | 1053 | ACGTAAATA | 1062 |

RESULT 6
ABL32517/c
ID ABL32517 standard; DNA; 5930 BP

AC ABL32517;

DT 26-MAR-2002 (First entry)

DE Human immune system associated gene SEQ ID NO: 490

KM Human/immune system disease, cytosine methylation, antiesthematic
KM antipertussic, antianemic, cytosolic, nocropic,
KM antipertussic, antianemic, cytosolic, nocropic,
KM neuroprotective, anti-HIV, anticonvulsant, ophthalmologic,
KM antihemmatic, antithyritic, antidiabetic, antipsoriatic,
KM antiinflammatory, cancer, eye disease, arteriosclerosis, anaemia,
KM acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy,
KM neurofibromatosis, rheumatoid arthritis, psoriasis, bowel disease,
KM gene, ds.

OS Homo sapiens.

PN WO200200928-A2

PD 03-JAN-2002.

PF 02-JUL-2001; 2001MO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPIGENOMICS AG

PI Olek A, Piepenbrock C, Berlin K,

DR WPI; 2002-130909/17

PT Nucleic acid comprising fragment of chemically modified gene, useful for diagnosis and treatment of diseases associated with abnormal

PS Claim 1; SEQ ID NO 490; 32pp + Sequence Listing; German

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/allergic bowel diseases. The present sequence is a gene of the invention.

Sequence 5930 BP; 2122 A; 24 C; 912 G; 2872 T; 0 other;

| | | | | |
|-----------------------|-----------------|--------------------|-----------|--------------|
| Query Match | 4.6%; | Score 82.4; | DB 24; | Length 5930; |
| Best Local Similarity | 48.3%; | Pred. No. 2.7e-05; | | |
| Matches 230; | Conservative 0; | Mismatches 246; | Indels 0; | Gaps 0 |

Qy 936 AAAATTGTAATCATACACTTTATTAATGTGTCCTATTTAAATTAAGTAAGTGTAT 995
Db 5149 AAAAAATAAAAAATATATATATAAACAATTACATCTTAAATATTTATAAACAATATTTT 509
Qy 996 CATTAAGATTGGAAGAAACCAATAGTCTCGTCTGATTTTGAATTAATTTGTTTGTATG 105
Db 5089 TAAAAAATTTACTATCAACTATATCAAAACCTCAACTGCCAATTTATATCTTAAT 503
Qy 1056 TTACTTTTCTCAAGCCTATATAAAACTTTGTACTTAATTTGATGCTGAAAAAAA 111

Db 5029 TTATATATTTTCACAAACATATTTAAATTAACAAAAATACAAATTTAAATTAATTAATTAAT 4970
QY 1116 TGTGTAATGAAATTCATAGAAATTAATGTAATTTAAAGTCCAAATTCATCATAGAAAT 1175
Db 4969 ACAATATATATTTACTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4910
QY 1176 TTAGTACAAAACGTAAGTCAAAAATATCTCTTAATTTTAAATTTTACACATATTAATTAAT 1235
Db 4909 AACAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4850
QY 1236 TATTCCTTAATTTAAATTTTACATATATTAATTTATCCCTGCTCCTTTGAAATACC 1295
Db 4849 AACTCTAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4790
QY 1296 ACCAACAATATTAATTAATTAATTTTATTTTATTAATTTTGAATCTCCTCAATATAT 1355
Db 4789 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4730
QY 1356 CTGATATTTATTTATTTTGTGTCATATTTTCTTATGTTTGAAGTTAACCTTTA 1411
Db 4729 ATTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4674

RESULT 7

ABL34492
ID ABL34492 standard; DNA; 11996 BP.

XX ABL34492;
AC
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEO ID NO: 45.

KM Metastasis associated gene; cytostatic; gene therapy; cancer;
KW cytosine methylation; gene; ds.

XX Homo sapiens.

XX MO200177376-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP03970.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPiG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-010922/01.

XX New nucleic acid derived from chemically treated metastasis genes,
PT useful for diagnosis of cancers by analysis of cytosine methylation,
PT also for treatment

XX Claim 1, SEQ ID NO 45; 23bp + Sequence Listing; English.

XX The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention.

XX Sequence 11996 BP; 2961 A; 261 C; 3028 G; 5746 T; 0 other;

XX Query Match 4.5%; Score 80; DB 24; Length 11996;

XX Best Local Similarity 49.0%; Pred. No. 7.1e-05;

XX Matches 242; Conservative 0; Mismatches 250; Indels 2; Gaps 1;

QY 912 ACTACTTAGTAGTATTAAGAAATTTGATATCATCATTTTAATTTATGT-CCT 969

Db 929 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 988
QY 970 TATTTAAATTTATGTAAGGTTGATTCATTTAGATTTGAGAAAACCAATAGTCTGCTCT 1029
Db 989 TATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1048
QY 1030 TGAATTTGAATTTATTTGTTTCTATGTTACTTTTCTCAAGCTATTTAAACCTTTGTA 1089
Db 1049 TATATTTATTTAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1108
QY 1090 ATGCTAAATTTGATGCTGGAATAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1149
Db 1109 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1168
QY 1150 AAAGTCCAAATTCCTCAATTTAGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTTAG 1209
Db 1169 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1228
QY 1210 TTTTAAATTTTACCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1269
Db 1229 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1288
QY 1270 TTATGACCTGTCACCTTTAGAAATTTACCAATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1329
Db 1289 TATATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1348
QY 1330 AATATTTTGAAGATCTCCCAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1389
Db 1349 ATTATTTATTTGTTGATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1408
QY 1390 TATGTTTATGAGTT 1403
Db 1409 GATGAGTTTGT 1422

RESULT 8

AAH93026/C
ID AAH93026 standard; DNA; 700 BP.

XX AAH93026;

DT 09-OCT-2001 (first entry)

XX Human inflammatory bowel disease related gene fragment ICR3119a.

XX Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;

XX single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;

XX chromosome 5q11-33; forensic test; gene therapy; ds.

XX Homo sapiens.

XX WO200142511-A2.

PD 14-JUN-2001.

PF 11-DEC-2000; 2000WO-US33632.

PR 10-DEC-1999; 99US-0170257.

PR 10-APR-2000; 2000US-0196046.

XX (WHEB) WHITEHEAD INST BIOMEDICAL RES.

XX (ELLIS) ELLISIS BIOTHERAPEUTICS CORP.

XX Daily M, Hudson TJ, Lander ES, Rioux J, Siminovitch K;

XX WPI; 2001-367874/38.

XX Testing for the presence of polymorphisms associated with inflammatory
PT bowel disease, using a hybridization assay -
XX Disclosure; Page 384; 463bp; English.

CC The present invention describes a method for detecting the presence of
 CC polymorphisms associated with inflammatory bowel diseases such as
 CC ulcerative colitis and Crohn's disease. The methods can be used to detect
 CC the presence of genetic polymorphisms associated with inflammatory bowel
 CC disease and correlating their occurrence with disease states. They may be
 CC used in this way for phenotypic correlations, forensics, paternity
 CC testing, medicine and genetic analysis. The present sequence is a gene
 CC containing a polymorphic site described in the exemplification of the
 CC invention.

XX Sequence 700 BP; 370 A; 13 C; 14 G; 302 T; 1 other;

Query Match 4.5%; Score 79.8; DB 22; Length 700;

Best Local Similarity 48.5%; Pred. No. 6.6e-05;

Matches 315; Conservative 0; Mismatches 323; Indels 12; Gaps 3;

QY 919 TTAGTAAAGTTGATGATTAAGTAAATGTAATCAGCATTTATTTTGTCTTATTAAA 978
 DB 697 TTAAAT 638
 QY 979 TTATGATAAAGTTGATGATTAAGTAAATGTAATCAGCATTTATTTTGTCTTATT 1038
 DB 637 TTAAAT 578
 QY 1039 TATATATGTTTCTGATGATTTCTTCAAGCTATATATATATATATATATATAT 1098
 DB 577 TAAAT 518
 QY 1099 TGTATGCTGAAAAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1158
 DB 517 TAAAT 458
 QY 1159 TATCATCATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 1218
 DB 457 TAT-----TAT 403
 QY 1219 TTAACAACAAT 1278
 DB 402 TAT 343
 QY 1279 GTACACCTTGAATGCCCAACAATATATATATATATATATATATATATATAT 1338
 DB 342 TAT 285
 QY 1339 GAGATCTCTCAAT 1398
 DB 284 TAT 225
 QY 1399 GAGTAAACCTTATATCTGCTCAACTAGTAATCAATATATATATATATATAT 1458
 DB 224 TATTTA-----TATATATATATATATATATATATATATATATATATATAT 170
 QY 1459 ATTGACATCTGAAACATATGTTTAACTGTTGAAATGTTAAAGTAATAAAGATTC 1518
 DB 169 AAT 110
 QY 1519 AGAATTAAGACCAT 1568
 DB 109 ATATTTAT 60

RESULT 9
 ABL34493/C
 ID ABL34493 standard; DNA; 11996 BP.

XX ABL34493;
 XX 26-MAR-2002 (first entry)
 XX Human metastasis associated gene SEQ ID NO: 46.
 XX Metastasis associated gene; cytostatic; gene therapy; cancer;
 XX "cytosine methylation; gene; ds."
 KW

XX Homo sapiens.
 OS
 XX
 XX WO200177376-A2.
 FN
 XX
 XX 18-OCT-2001.
 PD
 XX
 XX

XX 06-APR-2001; 2001WO-EP03970.
 PF

XX 06-APR-2000; 2000DE-1019058.
 PR

XX 07-APR-2000; 2000DE-1019173.
 PR

XX 30-JUN-2000; 2000DE-1032529.
 PR

XX 01-SEP-2000; 2000DE-1043826.
 PR

XX (EPIC-) EPIGENOMICS AG.
 PA

XX Olek A. Piepenbrock C. Berlin K;
 PI

XX WPI; 2002-010922/01.
 DR

XX New nucleic acid derived from chemically treated metastasis genes,
 PT useful for diagnosis of cancers by analysis of cytosine methylation,
 PT also for treatment.

XX Claim 1; SEQ ID NO 46; 23bp + Sequence Listing; English.
 PS

XX The present invention provides a number of human metastasis associated
 CC genes which are modified by cytosine methylation. The sequences can be
 CC used in the diagnosis and treatment of cancer. The present sequence is
 CC one of the genes of the invention.

XX Sequence 11996 BP; 2966 A; 261 C; 3041 G; 5728 T; 0 other;
 SQ

Query Match 4.4%; Score 77.8; DB 24; Length 11996;
 Best Local Similarity 50.1%; Pred. No. 0.00017;
 Matches 248; Conservative 0; Mismatches 242; Indels 5; Gaps 2;

QY 912 ACTACCTTGAAGTAT 969
 DB 11068 AT 11009
 QY 970 TATTTAAAT 1029
 DB 11008 TATTTAAAT 10949
 QY 1030 TGAATTTGAAT 1089
 DB 10948 TAAATTTAT 10889
 QY 1090 ATGCTAAAT 1149
 DB 10888 AT 10829
 QY 1150 AAAGTCCAAATCATCAATATATATATATATATATATATATATATATATATATAT 1209
 DB 10828 AT 10772
 QY 1210 TTTTAAATTTTAAACAAATATATATATATATATATATATATATATATATATAT 1269
 DB 10771 AT 10712
 QY 1270 TTAATACCTGCACTTGAATATATATATATATATATATATATATATATATATAT 1329
 DB 10711 AT 10652
 QY 1330 AATATATTTGAGATCTCTCAATATATATATATATATATATATATATATATATAT 1389
 DB 10651 AT 10592
 QY 1390 TATGTTTGAAGTTA 1404
 DB 10591 TAAATATATATCTTA 10577

RESULT 10
 ABL54336
 ID ABL54336 standard; DNA; 6641 BP.
 XX
 AC ABL54336;
 XX
 DT 29-JUL-2002 (first entry)
 XX
 DE Chemically treated apoptosis gene complementary to gene #18.
 XX
 KW Apoptosis; HIV; Bloom syndrome; cardiopathy;
 KW neurodegenerative disorder; Herpes simplex virus;
 KW amyotrophic lateral sclerosis; cancer; ds.
 XX
 OS Unidentified.
 XX
 PN WO200177164-A2.
 XX
 PD 18-OCT-2001.
 XX
 PE 06-APR-2001; 2001WO-EP03969.
 XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPig-) EPIGENOMICS AG.
 XX
 PI Olex A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-017444/02.
 XX
 PT Chemically modified sequences of genes associated with apoptosis are
 useful to determine methylation patterns of genomic DNA samples for
 PT diagnosis of associated diseases such as cancer -
 XX
 PS Claim 1; Seq ID #36; 24bp; English.
 XX
 CC This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented in the printed specification but is based on
 CC information supplied by the European patent office.
 XX
 SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;
 Query Match 4.3%; Score 76.4; DB 24; Length 6641;
 Best Local Similarity 45.9%; Pred. No. 0.00028;
 Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;
 QY 755 AAGTGGCACTTCCCTTGTGAGCTTCATAGAGCACTTACAGACCTTTTAACTAAAT 814
 Db 2561 AT 2620
 QY 815 ACTACAAAGAGAGAGATTTTAAACACTTGAGAGATTAAGGAGTTAAAGACACACAT 874
 Db 2621 ATATTTAAAGTAATTAATTTATTTTAAATATAGAGATATGATTTAAAGTGTGATG 2680
 QY 875 TAAGGGGAGGTGTTAAATTAATGCTGTAAACACACTTACCTTATAGATTAATA 934
 Db 2681 TGAGAAAGTTAGATATTTTATTTTGTATTTTAAAGATATATTTAGTTTAT 2740
 QY 935 GAAATTTGTAATATACATTAATTAATTTGCTTATTTAAATTAATATAGTAAAGTTGA 994
 Db 2741 TTAAATATTAAGATAGTTTATATATATATATTTTAAAGAAATATGTTAGATAT 2800

QY 995 TCATTAGATTAGAGAAACCAATAGTCTCGTCTGATTTTGAATTAATGTTTCAT 1054
 Db 2801 TGAATTAATTTGGTTTGAAGATATGATTTATTTTATTTAGCTATTTTGGATTATAT 2860
 QY 1055 GTTACTTTTCTTCAGACCTATATAAACCCTTTGA--ATGCTAATTTGATCGGGAA 1111
 Db 2861 GTTATATATGAATTTATTTATTTTATTTAGAAAGATATATGAAATTAATTAAGATATA 2920
 QY 1112 AAAATGTATAGTAATTCATAGAAATTAATG--TATTTCAAGTCCAAATCCATCAAT 1169
 Db 2921 GTTATATTAATAAATGTAATTTTATTTTATGATTTTGTGATTAATAAATGAAATTTAT 2980
 QY 1170 AAAAAATTTGTCACAAAGCTACTAAAAATTTCTCTATTTTAAATTTTCAACATA 1229
 Db 2981 ATTAAGTAATGATTAATTTATTTATATATATGATTAATTAATTAATTAATTAATTA 3040
 QY 1230 TAAAAATATCTCTATTTTAAATTTTCAATATATATATTAATACCGTCACCTTTAG 1289
 Db 3041 GTATTTATTTATTTATATATGATTTTAAATTTAGTAATTTATTTATTTATATATA 3100
 QY 1290 AATACACCAACAATTAATTAATCTTATATTTATTTCTTAATTAATTTGAGATCTCA 1349
 Db 3101 ATTAGTAATTTATTTATTTATTTATGATTAATTAATTAATTAATTTATTTATGATAT 3160
 QY 1350 ATATATCGATTTATTTATTTA--TATTTGTCATTTTCTATGTTTGAAGTTAAC 1407
 Db 3161 AATTAATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3220
 QY 1408 CTATATCTTGTGTCACAGTATTAATTCATATATATGATTTTGAAGACACATGACATC 1467
 Db 3221 ATTAAGTTTATTTTAAAGATTTTAAATATGATATATTTGATTTTATATATAATTA 3280
 QY 1469 TTGAAACATGTTTAACTTTGATGATGTTAAAGTAAAGTAAACATTCGAATTTATG 1527
 Db 3281 GTAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3340
 QY 1528 ACCATCTATTAATTAATCTCTGCTTGTCTT 1557
 Db 3341 TATATGATTTAGGTTTAAATTTTATTTT 3370

RESULT 11
 ABL32315
 ID ABL32315 standard; DNA; 6641 BP.
 XX
 AC ABL32315;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 288.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cyrostatic; noctropical;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; anticanceritic; antidiabetic; antipsozotic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PE 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPig-) EPIGENOMICS AG.
 XX

PI Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -

XX PS Claim 1; SEQ ID NO 288; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;

XX Query Match 4.3%; Score 76.4; DB 24; Length 6641;

XX Best Local Similarity 45.8%; Pred. No. 0.00028;

XX Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;

QY 755 AATGTCGACGTTCCTTCGCGAGTTCATAGACACTTACGACACTTTACTCTAAAT 814
DB 2561 AT 2620
QY 815 ACTACAAAGAGAGAAATTTTACAACTTAGAAGTAATGGAGTTAAAGCACACAT 874
DB 2621 ATATTAAGTAATTAATTTATTTTAAATTAAGAGATATGATTAAGAGTTGTATG 2680
QY 875 TAAAGGAGAGTTAAATTAATGTTGTTTACACACACTTACCTTTAGTAAATATA 934
DB 2681 TGGAAAGTTAGATATTTTATTTTGTAAATTTTAAAGATATATTTAGTTAT 2740
QY 935 GAAATTTGATATCATCATATATATATGCTCTTATTAATTAATGATAAGTTGTA 994
DB 2741 TTAAATTTAAGATAGTTTATTTATTAATATATTTTAAAGAAATATGTTAGAT 2800
QY 995 TCATTAAGATTGAGAAAACCAATATGCTCGTCTGATTTTGAATTAATGTTTCTAT 1054
DB 2801 TGAATTAATTTTGTAGAGATTTAGATTTTATTTTGTAGTTATTTTGTAGTTATAT 2860
QY 1055 GTTACTTTTCTCAAGCTATATTAATTAACCTTGA---ATGCTAAATTTGATCTGAAA 1111
DB 2861 GTTATATATGATTAATTTATTTATTTAGAAATATATGAACTTATTAATAA 2920
QY 1112 AAAATGTATATGATTAATTAATTAATG--TATTCGAAAGTCCAAATCCATCAT 1169
DB 2921 GTTATATTTTAAATATGTAATTTTATTTTGTATTTTGTGATTAATAATGAAATTTAT 2980
QY 1170 AGAAATTTAGTACAAAGCTTACCAAAATCTCTTATTTTAAATTTTACACATA 1229
DB 2981 ATTAAGATATGATTAATTTATTTATTTAGATTAATTTATTTATTTTAAATATTA 3040
QY 1230 TAAAAATTTCTTATTTTAAATTTTACATTAATTAATTTTACACCTGCACTTTAG 1289
DB 3041 GTTATTTATTTATTTATTTATTTATTTAGTATTTTATTTATTTATTTATTAATA 3100
QY 1290 AATACCAACAATATTTATTTACTAGATTTTATTTCTTATTAATTTTGTAGATCTCA 1349
DB 3101 ATTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3160
QY 1390 ATATATCTGATTTATTTATTTA--TATTTGTCATTTTCTTATTTTGTAGAGTTACC 1407
DB 3161 AATTAATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 3220
QY 1408 CTATATCTTGTCGCAACTAGTAAATCAATATGAGTTGTGAAGGACATTTGACATC 1467
DB 3221 AATAAGTTTATTTTAAATTTTATTAATAATGATGATTTATTTATTTATTTATTTATTTA 3280

QY 1468 TTGAACATTTGTTTACCTGTTGATGTTAAAGATATATAAATCAGATTTATG 1527
DB 3281 GTAGAAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAAT 3340
QY 1528 ACCATCTATTAATATTTACTTCTTTGCTTT 1557
DB 3341 TATATGAATTTAGGTTTAAATTTTATTTT 3370

XX RESULT 12

XX ABL32254
XX ID ABL32254 standard; DNA; 18154 BP.

XX ABL32254;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 227.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antianaemic; cytosine; noctropic;

XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX antineumatic; antiarthritic; antidiabetic; antipsoriatic;

XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

XX for diagnosis and treatment of diseases associated with abnormal

XX cytosine methylation -

XX PS Claim 1; SEQ ID NO 227; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

XX CC genes which are modified by the methylation of cytosines. The sequences

XX CC can be used in the diagnosis and treatment of immune system disorders,

XX CC including eye diseases such as retinopathy, neovascular glaucoma and

XX CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

XX CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

XX CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

XX CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 18154 BP; 5274 A; 112 C; 3279 G; 9489 T; 0 other;

XX Query Match 4.3%; Score 76.4; DB 24; Length 18154;

XX Best Local Similarity 47.2%; Pred. No. 0.0003;

XX Matches 303; Conservative 0; Mismatches 331; Indels 8; Gaps 2;

QY 927 TATTAAGAAAATTTGTAATCATCATTAATTTATTTGCTCTTATTTAAATTTATGTA 986
DB 13819 TTTTATTAATTAATTTTGTATATTTATTTTATTTTAAATTTATTTTATTTGCTTT 13878
QY 987 AAGTGTATCATTAAGATTGAGAAAACCAATATGCTCGTCTTATTTGAATTTATG 1046
DB 13879 TAAATTTATTTATTTGA--AATTTATTTAGTATTTGTTATGATTTTAAATTTATTTATTT 13937

| | | | |
|-----------|---|---|-------|
| Qy | 104 | TTTTGATGCTACCTTTCTTCAAGCCATATATAAAACCTTGTATGCTAAATGATGCT | 1106 |
| Db | 13938 | TATTTTTTTTATTTTAAATAGATATGTGATTAATTTTTTTTTTAAATAGTTTTAT | 13997 |
| Qy | 1107 | GGAAAAAATGTGATAGATTCATATAGAAATATGTGATTTCCAAAGTCCAAATCCATC | 1166 |
| Db | 13998 | TGATTTGATGTTTTTACGTTTTTTAAAAAATATTTTGAATGTGATATTTTTTTATG | 1405 |
| Qy | 1167 | AATAGAAATTTAGTCAAAACCGTACGCAAAATATCTCTATTTTAAATTTTACACA | 1226 |
| Db | 14058 | TTTTTTATTTTATTTTGTGTTTTTATTTTATATAATTTTGATTTTTTGAAGTTTTG----- | 14112 |
| Qy | 1227 | AATATAAAATATCTCTATTTTAAATTTTACATATAATATAATTTATCAGCTGCACCTT | 1286 |
| Db | 14113 | --ATTAGAAATATATTTATTTATTAATTTTTTTAAAGTAAAGATGATTTATTTTATAA | 14170 |
| Qy | 1287 | TGGAATACCACCAACAATATTAATCTAGATATTTTATCTTAATATTTTGAGATCTC | 1346 |
| Db | 14171 | TAAATTTTGAAGATTTTTTTTGAAGTATTAACATTTTATATAATTTATATATAT | 14230 |
| Qy | 1347 | TCAATATATCTGATATTTATTTATTTGTCGCAATTTTCTATAGTTTGAGTTAAC | 1406 |
| Db | 14231 | TTTAAATATTTTTTATTTTGTGTATGATTTTATTTTAAATTTATTTGTATATGT | 14290 |
| Qy | 1407 | CCTTATATCTTGTGCACAACGTATATCAATATATGAGTTTGGAAGACACATTTGACAT | 1466 |
| Db | 14291 | ATTTAAAGTTGTAAATATATATATTTTATGTGATATATGATGTGTGTTTTATTAAGT | 14350 |
| Qy | 1467 | CTTGAAACATGTGTTTATACCTGTGTGAATGTAAAGTAAATAAACATTCGAAATAT | 1526 |
| Db | 14351 | TTTATTTTATTTAGTTTTTATTTATTTATTTGATTTGTGAAATATATTTTAAAGATATATA | 14410 |
| Qy | 1527 | GACCATCTATTAATATCTCTCTTTGCTTTTAAAAAGTGT | 1568 |
| Db | 14411 | TAAATATAGAAATATATATAGTTTTTATTTGTGAAATTTTTT | 14452 |
| RESULT 13 | | | |
| ABZ10015 | | | |
| ID | ABZ10015 | standard; DNA; 19634 BP. | |
| XX | ABZ10015; | | |
| AC | | | |
| XX | 16-JAN-2003 (first entry) | | |
| DT | | | |
| XX | | | |
| DE | Haematopoietic cell proliferation disorder related DNA sequence #155. | | |
| XX | | | |
| KW | Human; haematopoietic cell proliferation disorder; cytostatic; gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia; cytosine methylation state; gene; ds. | | |
| XX | | | |
| OS | Homo sapiens. | | |
| XX | | | |
| PN | MO2002727272-A2. | | |
| XX | | | |
| BD | 03-OCT-2002. | | |
| PF | 26-MAR-2002; 2002WMO-EP03401. | | |
| XX | | | |
| PR | 26-MAR-2001; 2001US-278333P. | | |
| PA | (EPIG-) EPIGENOMICS AG. | | |
| XX | | | |
| FI | Berlin K, Braun A, Distler J, Gietig D, Howe A, Mueller J, Olek A, Piepenbrock C, Adorjan P, Grabs G, Ilesche R, Leu E, Levin A, Lipscher E, Mater S, Model F, Mueller V, Otto T, Pellet C, Schwabe I, Ziebarth H; | | |
| XX | | | |
| DR | WPI; 2003-018942/01. | | |
| XX | | | |
| PT | Detecting and differentiating between hematopoietic cell proliferative disorders, comprises contacting a target nucleic acid with a reagent | | |

| | |
|---------------------------|---|
| PT | that distinguishes between methylated and non-methylated CpG dinucleotides |
| PS | Claim 28; SEQ ID 155; 117bp; English. |
| XX | |
| CC | The present invention describes a method for detecting and differentiating between haematopoietic cell proliferative disorders associated with at least 1 gene and/or their regulatory regions in a subject. The method comprises contacting a target nucleic acid in a biological sample obtained from the subject with at least 1 reagent, which distinguishes between methylated and non-methylated CpG dinucleotides within the target nucleic acid. ABZ09861 to ABZ11118 represent specifically claimed nucleotide sequences from the present invention. Oligonucleotides from the present invention can be used: for differentiating between healthy hematopoietic cells and proliferative disorder hematopoietic cells; for differentiating between acute lymphocytic leukaemia and acute myelogenous leukaemia; as probes for determining the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) of haematopoietic cell proliferation disorder related sequences and their complements; and as primers for the amplification of haematopoietic cell proliferation disorder related DNA sequences. The nucleotide sequences from the present invention can also be used for detecting a predisposition to, or differentiation between subclasses, diagnosis, prognosis, treatment and/or monitoring of haematopoietic cell proliferative disorders. The present method enables a highly specific classification of haematopoietic cell proliferative disorders allowing for improved and informed treatment of patients. |
| SQ | Sequence 19634 BP; 5528 A; 272 C; 3979 G; 9855 T; 0 other; |
| Query Match | 4.3%; Score 76.2; DB 25; Length 19634; |
| Best Local Similarity | 45.4%; Pred. No. 0.00033; |
| Matches 314; Conservative | 0; Mismatches 373; Indels 4; Gaps 1 |
| OY | 874 TTAAAGGGGACGTAAATTAATTAATGTTTGNACACACACTTCCTTAGTAAGTATTATA 933 |
| Db | 6242 TTAAAGTTTTAATGTTAATTAATTTGGTAGAATTAAATTTGAATGAGTAATAGTTTAAIT 6301 |
| OY | 934 AGAAATTTGTATCATCATTAATTAATTAATGTCCTTAATTAATTAATTAATGATAAGTGTT 993 |
| Db | 6302 TAGTGAAGGTTTTTTTTTTTTTTTCGTTTTTTGGATTTGAAAAATTTTTTTTTTTTTTA 6366 |
| OY | 994 ATCATTAAGAATTGACAAAACAATAGTCTCGTCTGATTTTGAATTATTTGTTTCTA 1055 |
| Db | 6362 ATTATATGATAGCAAAAAATTAATTCGTAAATATATTGGTATATTAATGAATATGA 6422 |
| OY | 1054 TGTTACTTTCTCAGGCCCTAATPAATAAAGCTTGTGAATGCMAATGTGATGCTGGAAAA 1113 |
| Db | 6422 TAGTTTTTTTTTTTATTAATTTGTTAGTAAGTTTGTAAATGTAATTAATTAAT 6481 |
| OY | 1114 AATGTGATGAATCAATAGAAATTAATGTAATTTCAAAGTCCAATAATCATCATATAGAA 1172 |
| Db | 6482 TTTTAAATTTAGTAATATTGAGATTAATTAATTTTAAATTGTTNGATGATGATTT 6541 |
| OY | 1174 ATTTAGTACAAAGCACTAACAAAATATCTCTATTTTAAATTTTACACATATATAA 1233 |
| Db | 6542 TAAATGATTAATAAGTTATATAGTGTTAAATTTAAGTTTAAATTAATGAAGTATGTAG 6601 |
| OY | 1234 AATATCTCTATTTTAAATTTTCAATATATATATTAATCACCTGCACCTTAGAATA 1293 |
| Db | 6602 ATTGTGAAGATTTATTAAGTATATATATTTGTTTAAATTTTTTTTAAATTTTTTTTT 6661 |
| OY | 1294 CCACCAACAATATTAAATCTAGATATTTTA---TTCTTAATAATTTTGAGATCTCTCA 1349 |
| Db | 6662 AAAATTTTGTGTTTGATTTTAAAGATTTTAAATAATTTTGTGATTTTATTTTAT 6721 |
| OY | 1350 ATATATCTGATATTTATTTATATTGTCACATTTTCTATATGTTTAAAGTAACCT 1409 |
| Db | 6722 ATTTTAAATTTTGTTATTTTATTTTGTTATTTTATTTTATTTTATTTTGGTATATTT 6781 |
| OY | 1410 TAATCTTGCTCAACTAGTAATCAATATATATGAGTTTGTGAAGACATATGACATCTT 1469 |
| Db | 6782 TTTTATTTTAAATATGATTAATAATTTTATTAATGATTAATGATGATGATTTTATTTATTT 6841 |

QY 1470 GAAACATGGTTTAACTGTGTAAGTAAAGGTAAATAAACATTCAGATTATGAC 1529
 DB 6842 TTAGTAAGATGTTTAAATTTTGTGTTTATTTTAAATTTTGTAGTTAATTT 6901
 QY 1530 CATCTATTAATACCTTCTTGTCTTTAA 1560
 DB 6902 TAAATTTTATTTTAAATGATTATGATA 6932

RESULT 14

ABL34123/c
 ID ABL34123 standard; DNA; 10329 BP.

AC ABL34123;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2096.

KW Human; immune system disease; cytosine methylation; antiaesthetic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 2096; 32pp + Sequence listing; German.

CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

CC Sequence 10329 BP; 2681 A; 56 C; 1865 G; 5727 T; 0 other;

Query Match 4.2%; Score 74.2; DB 24; Length 10329;

Best Local Similarity 47.8%; Pred. No. 0.00069;

Matches 313; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

QY 712 TTACATTAACACCACTTCATTGAGAGAGATTGAGAGAGAAATGTCAGTTACTT 771

DB 5962 TTCATTTTAAATCAATTAATGACATCTTATTAATCTTATTTATTTAAACAC 5923

QY 772 TCTGCAAGTTCAAGAGCACTTAAGACACTTTACTAA---ATACTAACAAAGAGA 827

DB 5922 TATCAATGCAATTAATTAATCATATATTCCTATTAATAAAAAACATTCATCTCTTAATTA 5863
 QY 828 AGATTTTAAACACTTAAGAAAGTAATGGAGTTAAAGAGCAACACTTAAGGGGAGGT 887
 DB 5862 AAAAAATTAATAAAAAAACAACAAAAATTAACAAATTAATTAATTAATTAACACA 5803
 QY 888 TAAATTAATGTTGTTGTAACACCACTACCTTTAGTATATTAAGAAATTTGTAATC 947
 DB 5862 TTATTTTATTAACCTACCACTTAAATTAATTAATTAATTAATTAATTAATTAATTA 5743
 QY 948 ATCACTATTAATTAATTTGCTTATTTAAATTAATTAATTAATTAATTAATTAATTA 1007
 DB 5742 ACCCTATTTACAAAAAATCTATTTTAAATTAATTAATTAATTAATTAATTAATTA 5683
 QY 1008 GAAACCAAAATAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1067
 DB 5682 TATTAATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 5623
 QY 1068 AAGCTATTAATAAACTTTGTAAT-GCTAAATGTAAGCTGGAATAAAATGTAATGA 1126
 DB 5622 AAACATAAAATTAACCTTTTCATTAACAACTATTAATTAATTAATTAATTAATTAAT 5563
 QY 1127 TTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1186
 DB 5562 TTAATAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5503
 QY 1187 GGTAACTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1246
 DB 5502 CAATATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5447
 QY 1247 TTTAATTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1306
 DB 5446 AATAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5387
 QY 1307 TAACTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1361
 DB 5386 ATATATTTCCAAAAAACAATCAAAAAAATTAATTAATTAATTAATTAATTAATTAAT 5332

RESULT 15

ABL33013

ID ABL33013 standard; DNA; 7597 BP.

AC ABL33013;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 986.

KW Human; immune system disease; cytosine methylation; antiaesthetic;
 KW antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

DR WPI: 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 986; 32bp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 7597 BP; 1965 A; 109 C; 1790 G; 373 T; 0 other;

Query Match 4.28; Score 74; DB 24; Length 7597;
Best Local Similarity 43.88; Pred. No. 0.00074;

Matches 367; Conservative 0; Mismatches 470; Indels 1; Gaps 1;

QY 938 AATTGATCATCATTAATTAATTTGCTTATTTAAATTAATGATAAAGTTGATCA 997
Db 6 AAAGGTATGAGATATATGATATTTTGGTTTGTATGTTGGGAGAGAGGTTTAG 65
QY 998 TTAAAGATTGAGAAACCAATAGCCGCTGCTGATTTTGAATTAATGTTCTATGTT 1057
Db 66 TTTGGGGTTTATTAATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATG 125
QY 1058 ACTTTTCTTCACGCCCTATATAAACTTGTAAATGCTAAATGATGCTGAGAAAAATG 1117
Db 126 TATTTTATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 185
QY 1118 TGTAAATCATCAATAGAAATTAATGATTAATTAATTAATTAATTAATTAATTA 1177
Db 186 TTTTATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 245
QY 1178 AGTACAAAGCTACCTCAAAATATCTCTATTTTAAATTTACAAACAATATAAATA 1237
Db 246 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 305
QY 1238 TTCTCTAATTTAAATTTTACAATAATTAATTAATTAATTAATTAATTAATTAAT 1297
Db 306 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 365
QY 1298 CAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1357
Db 366 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 425
QY 1358 GATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1417
Db 426 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 485
QY 1418 GGTCAACTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1477
Db 486 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 545
QY 1478 GGTTTAACCCTGTGGAATGTTAAAGTAATTAATTAATTAATTAATTAATTAAT 1537
Db 546 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 605
QY 1538 AATAATCTTCCTTTGCTTTTAAAAAGGTGAGTAAAGTAAAGTAAAGTAAAGTAA 1597
Db 606 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 664
QY 1598 TGTCTGTGCGCGCTGTATATCAATTAATTAATTAATTAATTAATTAATTAATTAAT 1657
Db 665 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 724
QY 1658 TAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1717

Db 725 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 784
QY 1718 TCGGCTCTCTCTTTTCTTCTCAAGTATAAACAATGAATTAATTAATTAATTAATTAAT 1775
Db 785 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 842

Search completed: November 25, 2003, 14:33:59
Job time : 369 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 / Search time 4396 Seconds

(without alignments)
15592.781 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaactctgacatcttga.....tagagcgatcaagctgaacc 1783

Scoring table:

IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

Pred. No. is the number of results predicted by chance to have a

Score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| C 1 | 206 | 11.6 | 2282 | 8 | VFA277287 |
| C 2 | 101.4 | 5.7 | 4601 | 3 | DMU11584 |
| C 3 | 101.4 | 5.7 | 19517 | 3 | DMU37541 |
| C 4 | 94.2 | 5.3 | 25858 | 3 | AE014832 |
| C 5 | 93.6 | 5.2 | 137889 | 9 | AC073269 |
| C 6 | 91.8 | 5.1 | 8056 | 6 | AX599046 |
| C 7 | 91.4 | 5.1 | 2009 | 6 | AX457067 |
| C 8 | 91.4 | 5.1 | 8056 | 6 | AX599046 |
| C 9 | 90.8 | 5.1 | 155106 | 9 | AC104069 |
| C 10 | 90.2 | 5.1 | 250029 | 3 | AE014839 |
| C 11 | 90.2 | 5.1 | 176368 | 9 | CNS01RCG |
| C 12 | 90.2 | 5.1 | 181976 | 9 | AC131649 |
| C 13 | 90 | 5.0 | 104853 | 9 | AC117444 |
| C 14 | 90 | 5.0 | 150236 | 2 | BX248102 |
| C 15 | 90 | 5.0 | 153326 | 2 | BX119991 |
| C 16 | 89.8 | 5.0 | 108902 | 2 | AC011430 |
| C 17 | 89.6 | 5.0 | 118642 | 9 | AC126283 |
| C 18 | 89.6 | 5.0 | 130044 | 2 | AC137080 |
| C 19 | 89.4 | 5.0 | 30157 | 9 | AC073242 |
| C 20 | 89.4 | 5.0 | 94384 | 9 | AC011718 |
| C 21 | 89.4 | 5.0 | 106358 | 9 | ALB07813 |
| C 22 | 89.4 | 5.0 | 348600 | 1 | AB063521 |
| C 23 | 89.2 | 5.0 | 139409 | 9 | AC069181 |
| C 24 | 89.2 | 5.0 | 210797 | 9 | AC013391 |
| C 25 | 89 | 5.0 | 8056 | 6 | AX598900 |
| C 26 | 89 | 5.0 | 118642 | 9 | AC126283 |
| C 27 | 88.8 | 5.0 | 67970 | 3 | PFMA113 |
| C 28 | 88.8 | 5.0 | 152359 | 9 | AC093790 |
| C 29 | 88.8 | 5.0 | 159942 | 9 | AC025018 |
| C 30 | 88.4 | 5.0 | 104992 | 2 | AC005504 |
| C 31 | 88.4 | 5.0 | 138504 | 9 | AC092665 |
| C 32 | 88.4 | 5.0 | 159904 | 2 | BX322565 |
| C 33 | 88.4 | 5.0 | 169546 | 2 | AC004157 |
| C 34 | 88.4 | 5.0 | 250621 | 3 | AE014849 |
| C 35 | 88.4 | 5.0 | 349980 | 6 | AX44567 |
| C 36 | 88.2 | 4.9 | 1121 | 3 | AF489467 |
| C 37 | 88.2 | 4.9 | 14867 | 3 | AE001398 |
| C 38 | 87.8 | 4.9 | 349751 | 3 | PFMA1493 |
| C 39 | 87.2 | 4.9 | 141275 | 2 | BX510640 |
| C 40 | 87.2 | 4.9 | 313050 | 3 | PF5929352 |
| C 41 | 87 | 4.9 | 30157 | 9 | AC073242 |
| C 42 | 86.6 | 4.9 | 76568 | 3 | MBREV |
| C 43 | 86.6 | 4.9 | 152966 | 9 | AL359713 |
| C 44 | 85.8 | 4.8 | 104853 | 9 | AC117444 |
| C 45 | 85.8 | 4.8 | 110000 | 2 | PFMA113_03 |

ALIGNMENTS

RESULT 1
VFA277287/c VFA277287 2282 bp DNA linear PLN 07-DEC-2000
DEFINITION Vicia faba partial endodis gene, exon 1 and 5' UTR.
ACCESSION AJ277287
VERSION AJ277287.1 GI:11602752
KEYWORDS endodis gene.
SOURCE Vicia faba (fava bean)
ORGANISM Vicia faba
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
Vicia.
REFERENCE
AUTHORS Hohnjec,N., Kuster,H., Albus,U., Frosch,S.C., Becker,J.D.,


```

Db      2805 TATAGTTTTTTTTTAAAAAAAAAATTTTATTTTTTAAAAAATTTTTTTTAAAAATGAAA 2746
Qy      914 TACCTTAGTAGATATATAGAAAATTTGTAATCATACATTAATTAATTTGACCTTAT 973
Db      2745 AATTAATTAATTAATTAATTTTCATTAATAAATTTATTTATTTTATTTTATTTT 2686
Qy      974 TAAATTAATTAATTAATTTGTAATCATTAAGATGAGAAACCAATAGCTCCTCTGAT 1033
Db      2685 AAAAAACATGATTTTATTAATAAATATTTTTTTTAAAAAATTAACATTAAGAAAATTT 2626
Qy      1034 TTTTGATTAATTTGTTTCTATAGTACTTCTTCTTAACCCATTAATAAAACCTTGTAATG 1093
Db      2625 TAAAAATTTTATTAATTAATTTTAAAAAATTTTATTTTCTATTAATAATAATTA 2566
Qy      1094 TAAATTTGATGCGGAAAAAATGTAATGAATTCATTAAGAAATTAATGTAATTCAG 1153
Db      2565 TAAATTTTCATATAATATATAATTTAATAAATAATAATAATTAATTAATTAATAA 2506
Qy      1154 TCCAAATTCATCAATTAAGAATTTAGTACAAACGTACTCAAAATATTTCTTATTT 1213
Db      2505 AAAAAAATTAATTAATTTTATTTTATTAATTAATTAATTAATTAATTAATTAAT 2446
Qy      1214 AATTTTACAAATATTAATAAATATTTCTTATTTTAAATTTTACATTAATTAATTAAT 1273
Db      2445 TTAATTTAGTATTAATTAATTTATTTATTTATTAATTAATTAATTAATTAATTAAT 2386
Qy      1274 CACCTGACCTTAGAATACCAACCAATTAATTAATTAATTAATTAATTTATTTCTTA 1333
Db      2385 ATTTAATTAATTAATTAATTAATTTATTTATTAATTAATTAATTAATTAATTTATTT 2326
Qy      1334 ATTTTGATCTCTCAATATATCGATTAATTTATTTATTTGTCATTTTCTTATG 1393
Db      2325 ATATATATATATATATATATATATATATTTTATTAATTTATTTAA--ATATATTAATTA 2268
Qy      1394 TTTTAGATTACCTTATATCTTGTGCAACAGTATTAATTAATTAATTAATTAATTTGTAAG 1453
Db      2267 TAAAAATATTTATTAATAAATAGTTATTAAGATTAATTAATAATAATTTTTTTTA 2208
Qy      1454 GACCATGATCATTTAAACATTTGTTTAACCTGTTGGAATGTAAGTAATTA 1513
Db      2207 AAAAAAATATTTTAAATTTTAAATTAATAATAATTTAAGATGAGGGAATAAT 2148
Qy      1514 CATTCAGATTAATGACCATCTATTAATATATCTTCTTCTTTAAAGATGTCATG 1573
Db      2147 TTAATTTCAATTAATATATATATATATATATATATATATTAATTAATTAATTAATTA 2086
Qy      1574 A 1574
Db      2087 A 2087

```

```

RESULT 3
DMU37541/c 19517 bp DNA circular INV 02-MAR-2001
LOCUS Drosophila melanogaster complete mitochondrial genome.
DEFINITION U37541
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS
SOURCE
ORGANISM
TITLE mitochondrial Drosophila melanogaster (fruit fly)
JOURNAL Drosophila melanogaster
AUTHORS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phyloroidae; Drosophilidae; Drosophila.
REFERENCE 1 (bases 12511 to 12682)
AUTHORS Clary,D.O., Goddard,J.M., Martin,S.C., Fauron,C.M. and
Wolstenholme,D.R.
TITLE Drosophila mitochondrial DNA: a novel gene order
JOURNAL Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE 6294611
PUBMED 6294611
REFERENCE 2 (bases 5269 to 5695)
AUTHORS Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.

```

```

TITLE
JOURNAL Transfer RNA genes in Drosophila mitochondrial DNA: related 5'
MEDLINE flanking sequences and comparisons to mammalian mitochondrial rRNA
PUBMED genes
63220794 Nucleic Acids Res. 11 (8), 2411-2425 (1983)
3 (bases 404 to 5272) de Bruijn,M.H.
REFERENCE Drosophila melanogaster mitochondrial DNA, a novel organization and
AUTHORS genetic code
TITLE Nature 304 (5923), 234-241 (1983)
JOURNAL 83245048
MEDLINE 6408489
PUBMED 4 (bases 804 to 1778)
REFERENCE Satta,Y., Ishiwa,H. and Chigusa,S.I.
AUTHORS Analysis of nucleotide substitutions of mitochondrial DNAs in
TITLE Drosophila melanogaster and its sibling species
JOURNAL Mol. Biol. Evol. 4 (6), 638-650 (1987)
MEDLINE 88174373
PUBMED 2832697
REFERENCE 5 (bases 5268 to 13619)
AUTHORS Garesse,R.
TITLE Drosophila melanogaster mitochondrial DNA: gene organization and
JOURNAL evolutionary considerations
MEDLINE Genetics 118 (4), 649-663 (1988)
PUBMED 88212147
6 (bases 441 to 2967) Satta,Y. and Takahata,N.
REFERENCE Evolution of Drosophila mitochondrial DNA and the history of the
AUTHORS melanogaster subgroup
TITLE Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
JOURNAL 2124697
MEDLINE 91088557
PUBMED 7 (bases 14215 to 14512)
REFERENCE Ballard,J.W., Olsen,G.J., Fitch,D.P., Odgers,W.A., Rowell,D.M. and
AUTHORS Atkinson,P.W.
TITLE Evidence from 12S ribosomal RNA sequences that onychophorans are
JOURNAL modified arthropods
MEDLINE Science 258 (5086), 1345-1348 (1992)
PUBMED 14558257
8 (bases 14917 to 19517) Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
REFERENCE Sequence, organization, and evolution of the A+T region of
AUTHORS Drosophila melanogaster mitochondrial DNA
JOURNAL Mol. Biol. Evol. 11 (3), 523-538 (1994)
MEDLINE 94285822
PUBMED 8015445
9 (bases 1 to 408; 13319 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.
REFERENCE Drosophila melanogaster mitochondrial DNA: completion of the
AUTHORS nucleotide sequence and evolutionary comparisons
JOURNAL Insect Mol. Biol. 4 (4), 263-278 (1995)
MEDLINE 96423163
PUBMED 10 (bases 1 to 19517) Lewis,D.L., Farr,C.L. and Kaguni,L.S.
REFERENCE Direct Submission
AUTHORS Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
JOURNAL Michigan State University, East Lansing, MI 48824-1319, USA
FEATURES
source
1..19517
/organism="Drosophila melanogaster"
/organella="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="derived from new and previously submitted
sequences; sequence is a composite containing sequences
obtained from different Drosophila melanogaster strains"
1..65
/product="rRNA-11e"
complement(97..165)

```

tRNA
171..239
/product="tRNA-Gln"
/product="tRNA-Phe"
240..1265
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 2"
/protein_id="AAC47811.1"
/db_xref="GI:1166530"
/translation="MFNSSKILFTITMITIGLITVTSNMGAMWGEFINLSTPL
LSDNNIMSTFESIKYELTQVLAFTVLSSILMDKNNNEINESFTSMIMALL
LKSQAPFHEFPMNMEGLTMMALMTMOKIAPMLISYNIKYLISVLSVLI
GAIGCNOTSLRKLMASFSLNHLGMLSLMSISIMLIFEFYSLSEVLPMNFIF
KPLHNLQESFVNSKLIKPLTFLMNLISGGLPPFLGPLKMLVIOQLTLCQYEMLT
IMMSTLITLFFYLKICYSAPMNVEENNMKNNSINVMIMTFEFSIFGLFLI
SLRYFNP"
1264..1329
/product="tRNA-Trp"
complement(1322..1383)
/product="tRNA-Cys"
complement(1403..1468)
/product="tRNA-Tyr"
join(1470..1472,1474..3009)
/note="mechanism underlying reading frame shift after
first codon uncertain"
/codon_start=1
/transl_table=5
/product="cytochrome c oxidase subunit I"
/protein_id="AAC47812.2"
/db_xref="GI:7412843"
/translation="MSRQWLFSTNKKDGLTYIFGAMAGMGTSLIIRALBHPG
ALIGDOIVNVIVTAHAFIMFPMWPMIMGGGNLVPLMGADPMAPPRNNNSFW
LLPPLSLIVSMEVNGAGTGTVPPLSAGIAGHAGASVDLAFSLHAGLSLIGA
VNFRTYINRSTGISIDRMPFLFMSAVITALLLSLPLAGAITMLITDNLTSF
PDAGGDDPLTVOHLFMFEGHPEVLIILPGFMISHIISQSGKXETGSGMTYAM
LAIGLGRFYMAHMTVGMVDTPRAFTSATMTIIVPGIKIFSLATLHGTOLYS
PALMALGFVFLFTVGLTGVVANSVDIILHDTYVVAHRYVTSKAVVAINAGF
IHWPLEFGLTLNNKMLKSHFTIMFIGNVTFEPQPLGAGMPKRSYDPATYTN
IVSTIGSTLILGILFFFLIMESIVSOROVIVPQLNSIEVMYQTPAHSYSELPL
LVNN"
3012..3077
/product="tRNA-Leu"
3083..3367
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except=(pos:3767,aa:TERM)
/transl_table=5
/product="cytochrome c oxidase subunit II"
/protein_id="AAC47813.1"
/db_xref="GI:1166532"
/translation="MSTWANLGLQDSAPLMEOILFFPHDAIILVMTIVLYGMFK
LFENYVNRFLHGOILIMTWILPALILFLIALPSILRLYLDELINESVTLKSI
QWMSYVSDENNIEPDSYMIPTNELMTDFLLDNRVLYLPMNSOILIVTAADVI
HSMTVPALGVKVGDTGRLNQTNFIRPGLFYGCSEICGASHSEMPVIESVPNY
FIKXISSNNS"
3768..3838
/product="tRNA-Lys"
3840..3906
/product="tRNA-Asp"
3907..4068
/codon_start=1
/transl_table=5
/product="ATPase 8"
/protein_id="AAC47814.1"
/db_xref="GI:1166533"
/translation="MPQMAPISWLLFIIFSITFLFCSINYSYMPNSKSNELNMI
NLSNMWKM"
4062..4736
/codon_start=1
/transl_table=5
/product="ATPase 6"
/protein_id="AAC47815.1"

Query Match 5.7% Score 101.4; DB 3; Length 19517;
Best Local Similarity 46.5%; Pred. No. 2,4e-06;
Matches 363; Conservative 0; Mismatches 416; Indels 2; Gaps 1;
tRNA
794 TACAGACACTTTACTATAATACATACAAAGAGAGATTTAACACTTACAGAGTAT 853
Db TAGTA 17722
Qy GGGAGTTAAAGCAACACATTAGGGGAGAGTAAATTAAGTGTGTGAACACACAC 913
Db 17721 TATAGTTTTTTTAAAAAAAATATATTTTTTTTAAAAATTTTTTTTAAAAATGAAA 17662
Qy TACCTTTAGTAT 973
Db AATAAT 17602
Qy TAAATTTTGAATTAAGTGTATCATTTAGATTGAGAAAACCAATATAGCTCGCTGAT 1093
Db 974 TAAATTTTGAATTAAGTGTATCATTTAGATTGAGAAAACCAATATAGCTCGCTGAT 1093
17601 AAAAAACATGATTTAT 17442
Qy TTTGAATTTATGTTTCTATGTTACTTTCTTCAAGCCATATATATATATATATATATAT 1093
Db 1034 TTTGAATTTATGTTTCTATGTTACTTTCTTCAAGCCATATATATATATATATATATAT 1093
17541 TAAATTTTAT 17482
Qy TAAATTTATGCTGGAAAAAATGCTGATATCAATCAATCAATATATATATATATATATAT 1153
Db 1094 TAAATTTATGCTGGAAAAAATGCTGATATCAATCAATCAATATATATATATATATAT 1153
17481 TAAATTTATGCTGGAAAAAATGCTGATATCAATCAATCAATATATATATATATATAT 17422
Qy TCCAAAAATCCATCAATCAATCAATATATATATATATATATATATATATATATATATAT 1213
Db 1154 TCCAAAAATCCATCAATCAATCAATATATATATATATATATATATATATATATATAT 1213
17421 AAAAAAAT 17362
Qy AATTTTCAACAT 1273
Db 1214 AATTTTCAACAT 1273
17361 TTAATTTATGTTAT 17302
Qy CACCTGTACCTTTGAAATACCAACACATATATATATATATATATATATATATATATAT 1333
Db 1274 CACCTGTACCTTTGAAATACCAACACATATATATATATATATATATATATATATAT 1333


```
/db_xref="GI:23495048"
/translation="MRI FGLRWTYYSEYETYL CNEQSRKRFSTYMLYNTSEL
QNNMNLNNMNLNLSYGIKSVKGYVNMKPFHYNNLAKFYCNFTFEDIIILQ
SYALCERNEEYVLSLNKLVI FKMDEOEYNOQVNIYKYIIASOOLNDEYLI
TWELKIKNNLHYGLKISII LHSLTKEIYDOOLDSIYIYVNEBHKCNVSH
IISAVSKOINIKELQYKLOIYENIKYMSLSIYNLIIDKPIICVDYNNLSY
KIDODYCTONIKENKYSSEPDNLFOKKKKKDYDKNIYNESNVCDKIKYELON
NCSYNNNTYDHSNXYVHTYOSKETYPHTPIVDTYKAKTKCTTNINNDRKKK
NENNVNIFSNSSKHSNDEYCHLNDTLKINNNEONNLAKKKEKLIIONITPLFS
KNNVCIAFISFKQVLACTYKELNENYIYKLLPYLAKIKNNKISPEECVLI
ECPTVLPYIDNNIDEIINIILNLETVIFHYNYLCRIQCPQHLTIYNDLTSIDC
IIFKNKAERYCNLKDQLFLHFEYENDEQEMINYLKQLEONNNHIILOSSENSI
COOTHOHDSNTNMSNMNIYNDKLIYKYNKVOCFNEKEKNIYEATVNHDKXD
EMDGTPEVHNESNDYKXSDYIYENTINEOKNI"
repeat_region
complement (2821..2848)
/nc="AT_rich"
/rpt_type=tandem
complement (3335..3365)
/nc="AT_rich"
/rpt_type=tandem
complement (3453..3531)
/nc="AT_rich"
/rpt_type=tandem
complement (3861..3880)
/rpt_type=tandem
complement (4042..4111)
/nc="AT_rich"
/rpt_type=tandem
complement (4275..4301)
/nc="AT_rich"
/rpt_type=tandem
complement (4348..4369)
/nc="AT_rich"
/rpt_type=tandem
complement (4794..4992)
/nc="AT_rich"
/rpt_type=tandem
complement (4833..4872)
/rpt_type=tandem
complement (5154..5178)
/nc="AT_rich"
/rpt_type=tandem
complement (5275..5331)
/rpt_type=tandem
complement (5332..5357)
/nc="AT_rich"
/rpt_type=tandem
complement (5378..5402)
/nc="AT_rich"
/rpt_type=tandem
complement (5451..5484)
/rpt_type=tandem
complement (5495..5532)
/nc="AT_rich"
/rpt_type=tandem
complement (5578..5604)
/nc="AT_rich"
/rpt_type=tandem
complement (5670..5689)
/rpt_type=tandem
complement (5749..5838)
```

```
Query Match 5.3% Score 94.2; DB 3; Length 258658;
Beet Local Similarity 45.0%; Pred. No. 1.3e-05;
Matches 490; Conservative 0; Mismatches 583; Indels 16; Gaps 3;

repeat_region /rpt_type=tandem
repeat_region /rpt_type=tandem
complement (5876..5900)
/nc="AT_rich"
/rpt_type=tandem

360 TCTGACCTGTGCATATTTCTTACACAGTTAGAAAAATTTCTPAAGTGTGAAATTTGACT 419
187834 TATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187775

420 TTCCAAAGCAATTCACCTTTCACCTTCCTGATATAAACAACCTCATTTCTPAACATG 479
187774 TAAATTAATTTAATTAATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTA 187715

480 TCTTGATGAAATGTGATCTTGAAATTTGATGTGATGCAAAAGTCAAAGCTTGACTTT 539
187714 AAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187655

540 CAGTGTGCAATGCACTTTTGCCTGTGCGCAATTCCAACCTAATGTGATGATGAT 599
187654 ATATATTAATTAATTAATTTCTTTAATTTAATAAATAAATAAATAAATTAATTAATTAAT 187595

600 GCTGCA--AACTGATGCAATGGAAGATCTTATGAGAAAATCTTGAGACCTGAGAGAA 657
187554 AATAAATAATTTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187535

658 AAATTTGTGATGACACACAAAAGATCTGTTTTTCTAGTGGACAGACACTTACA 717
187534 TATTTACACAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187475

718 TAAACACCACTTCATTCGAAAGTGTGAGAGAAAATGTCAGTCACTTCTGCA 777
187474 AAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187415

778 GTTCATAGACCACTTACACACACTTTTACTMAAATACACAAAAGGAGATTTTAC 837
187414 TTTTAATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187355

838 AACTTGAAGAGTAAAGGAGTTAAAGACACACTTAAAGGGAGGTAAATTAAT 897
187354 AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 187295

898 GTGTTGTAACACACACTTACTTGAAGATTTAAAGAAATGTAAATCACTTAT 957
187294 ATATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 187235

958 AATTATTTGCTTATTTAAATTTATGATTAAGTGTGATCATTAAGTGAAGAAACCAA 1017
187234 TTTATTTATTTATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTTATTTAT 187175

1018 TAGTCCTGCTCTGATTTTGAATTAATTTGTTTCTAGTGTACTTTCTCAAGCCTAT 1077
187174 TTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187123

1078 AAAAAGCTTGAAGCTAAATTTGATGCTGAAAAAATGCTATGATTAATTCATAGAA 1137
187122 TTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187063

1138 TTATGCTAATTCAAAGTCAAAATTCATCAATAGAAATTTAGTCAAAAGCTCACTCAA 1197
187062 TATTAATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 187003

1198 AATATTTCTTATTTTAAATTTTACACAAATTAATTTCTTATTTTAAATTTTAA 1257
187002 AATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 186943

1258 CAATATTAATTTATTTACCTGTCACCTTGAATACCAACAAATTAATTAATTTA 1317
```

```

Db      186942 TTATTATTATTTT-----TATTATATTAAATAAATATTATTGATTATTATTATT 186889
Cy      1318 TATTATTCTTAAATATTGAGATCTTCAATATATTCGATATTATTATTATTGTT 1377
Db      186888 TATTATTATTATTATTATTATTATTAAATAAATAATTGATTATTATTATTATTA 186829
Cy      1378 GTCAATTCTTCTATGTTTAGAGTTAACCTTATATCTGGCAACTGTAATCAAT 1437
Db      186828 TTTTATTATTATTATTATTATTATTAAATAAATAATTATTGTTATTATTATTATT 186769
Cy      1438 ATATGAGTT 1446
Db      186768 TTATTATTTT 186760

RESULT 5
LOCUS   AC073269 137889 bp DNA linear PRI 29-APR-2003
DEFINITION Homo sapiens BAC clone RP11-436F9 from 7, complete sequence.
ACCESSION AC073269
VERSION   AC073269.7 GI:14269824
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiindae; Homo.
REFERENCE 1 (bases 1 to 137889)
AUTHORS   Sulston,J.E. and Waterston,R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED    9847074
REFERENCE 2 (bases 1 to 137889)
AUTHORS   Abbott,A., Mclellan,M. and Vanhurnt,A.
TITLE     The sequence of Homo sapiens BAC clone RP11-436F9
JOURNAL   Unpublished (2001)
REFERENCE 3 (bases 1 to 137889)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (12-JUN-2000) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 4 (bases 1 to 137889)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (31-MAY-2001) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 5 (bases 1 to 137889)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (01-JUN-2001) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE 6 (bases 1 to 137889)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (07-NOV-2001) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 137889)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 8 (bases 1 to 137889)
AUTHORS   Waterston,R.
TITLE     Direct Submission
JOURNAL   Submitted (29-APR-2003) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On May 31, 2001 this sequence version replaced gi:13794256.
          ----- Genome Center
          Center: Washington University Genome Sequencing Center

```

Center code: MUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0436609

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
 The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.V. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBac3.6

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP11-29C24, 2000 bp overlap the clone sequenced to the right is CTD-2016H9, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-436F9 actual end is at base position 92636 of CTD-2016H9.

FEATURES

```

source
1..137889
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7"
  /clone="RP11-436F9"
  /clone_lib="RPCT-11"
  453..773
    /rpt_family="L1"
    repeat_region
  774..1051
    /rpt_family="Alu"
    repeat_region
  1063..1316
    /rpt_family="L1"
    repeat_region
  1351..1469
    /rpt_family="L1"
    repeat_region
  2098..2256
    /rpt_family="MIR"
    repeat_region
  2505..2753
    /rpt_family="L1"
    repeat_region
  2971..3034
    /rpt_family="L1"
    repeat_region
  3358..3742
    /rpt_family="L2"
    repeat_region
  3905..4033
    /rpt_family="L2"
    repeat_region
  4148..5347
    repeat_region

```

```

variation      /rpt_family="L1"
                4179
                /allele="G"
                /allele="A"
                /db_xref="dbSNP:2177715"
variation      4590
                /allele="G"
                /allele="A"
                /db_xref="dbSNP:2177716"
repeat_region  5341..5560
                /rpt_family="L1"
repeat_region  6804..6903
                /rpt_family="L2"
repeat_region  6920..7957
                /rpt_family="MER1_type"
repeat_region  7958..8261
                /rpt_family="Alu"
repeat_region  8262..8370
                /rpt_family="MER1_type"
repeat_region  8833..8877
                /rpt_family="(TG)n"
repeat_region  8896..9211
                /rpt_family="L2"
repeat_region  9345..9387
                /rpt_family="L2"
repeat_region  10487..10723
                /rpt_family="MER2_type"
variation      11617
                /allele="C"
                /allele="T"
                /db_xref="dbSNP:1357887"
repeat_region  13743..14036
                /rpt_family="Alu"
repeat_region  14037..14078
                /rpt_family="AT-rich"
variation      14345
                /allele="G"
                /allele="C"
                /db_xref="dbSNP:1525148"
variation      14579
                /allele="G"
                /allele="A"
                /db_xref="dbSNP:1525147"
variation      14639
                /allele="C"
                /allele="A"
                /db_xref="dbSNP:1525146"
variation      14688
                /allele="T"
                /allele="C"
                /db_xref="dbSNP:1525145"
repeat_region  14918..15301
                /rpt_family="ERV1"
repeat_region  15333..15501
                /rpt_family="(TCRA)n"
variation      15492
                /allele="G"
                /allele="C"
                /db_xref="dbSNP:1525144"
repeat_region  15366..15748
                /rpt_family="(TCRA)n"
repeat_region  15764..15802
                /rpt_family="(TCRA)n"
repeat_region  15814..15980
                /rpt_family="(TCRA)n"
repeat_region  15980..16035
                /rpt_family="ERV1"
repeat_region  16236..16395
                /rpt_family="WALR"
variation      16520
                /allele="T"
                /allele="C"

```

```

Query Match      5.2% Score 93.6; DB 9; Length 137889;
Best Local Similarity 49.0%; Pred. No. 2e-05;
Matches 339; Conservative 0; Mismatches 344; Indels 9; Gaps 3;

QY 918 TTTACTAAGTATTAAGAAAATTGTAATCATCATTA-TAATTATGCTTAATTAA 976
DB TTTATTAATATTAATATATATATTAATATATATTTTATTAATATTAATATTAATTTCA 63108
QY 977 AAATATGATTAAGTATGATCTTAAGTGAAGAAACCAATAGTCCTGCTGATTTT 1036
DB TTTATTAATATTAATATATATATTAATATATATTTTATTAATATTAATATTAATTTCA 63168
QY 1037 TGAATTAATGTTTTCATAGTACTTTTCCTCAAGCCATATTAATAAATTGTAATGCTAA 1096
DB ATATATATTTATTAATATATTAATATTAATATTAATATTAATATTAATATTAATATTAAT 63228
QY 1097 ATTGATGCTGGAATAAATGTAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 1156
DB ATATTAATATTTTAATATTAATATTAATATTAATATTAATATTAATATTAATATTAATTTT 63288
QY 1157 AAAATCCATC-----AATGAAATTTAGTACAAAAGTAACTCAAAAATATTCCTTAAT 1210
DB ATTATATATGATATATATATATTTTAATATTAATATTAATATTAATATTTTATTAATATAT 63348
QY 1211 TTTAAATTTTACACATATATAAATATTCCTTAATTTTAAATTTTACAAATAATATAT 1270
DB TATATATATATTTTAAATATATATATTTTATTAATATTAATATTAATATTTTATATATAT 63408
QY 1271 TATCACTGTCACCTTTAGATATACCAACAATATTAATATTAATATTAATATTTATCTTA 1330
DB ATTATATATATATATATATATTAATATTAATATTAATATTAATATTAATATTAATATTAAT 63468
QY 1331 ATATATTTGAGATCTCAATATATATGATATTTATTTATTTATTTATTTATTTCTT 1390
DB ATA--TTTATATTTTATTAATATTAATATTAATATTAATATTAATATTAATATTTATTTAT 63526
QY 1391 ATGTTTGAATTAACCTTATATCTTGGTCAAGTAAATTAATTAATTAATTAATTAATTAAT 1450
DB ATATATATATATATATATATTAATATTAATATTAATATTAATATTAATATTAATATTAATTA 63586
QY 1451 AAGGACATTTGACATCTTAATTAATATGTTTAAACCTGTTGAATGTTAAAGTAAAT 1510
DB TATTAATATTAATATATTTTAACTTATTAATATTAATATTTTAACTTATTAATATTAATTA 63646
QY 1511 AAACATTCAGATTAATGACATCTTAATATTAATATCTTCCTTGTCTTAAAAAGTGTGC 1570
DB TATTTATATATTTTATATATATTAATATTAATATTAATATTAATATTTTATTAATATTAATAT 63706
QY 1571 ATGAATATGCTTAATGTAAGTAACTGCTCT 1602
DB TTTATATATATATATTTTATTAATATTAATAT 63738

RESULT 6
AXS99046/c 8056 bp DNA linear PAT 14-FEB-2003
LOCUS Definition Sequence 386 from Patent WO02077272.
ACCESSION AXS99046
VERSION AXS99046.1 GI:28399186
KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.
REFERENCE
1
AUTHORS
Berlin, K., Braun, A., Distler, J., Guetig, D., Howe, A., Mueller, J.,
Olek, A., Piepenbrock, C., Adorjan, P., Grabs, G., Leech, R., Liu, E.,
Levin, A., Lipscher, E., Maier, S., Model, F., Mueller, V., Otto, F.,
Pellet, C. and Ziebart, H.
Methods and nucleic acids for the analysis of hematopoietic cell
proliferative disorders
Patent: WO 02077272-A 386 03-OCT-2002;
JOURNAL
Epigenomics AG (DE)
FEATURES
Location/Qualifiers

```


Source 1. 8056
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 3711 a 0 c 371 g 3974 t

Query Match 5.1%; Score 91.8; DB 6; Length 8056;
 Best Local Similarity 45.0%; Pred. No. 8.8e-05;
 Matches 481; Conservative 0; Mismatches 572; Indels 16; Gaps 3;

QY 707 ACACATTAACATTAACACCCCTTCATTCGAGCGTATTGAGAGGAAATGTGACGTT 766
 DB 2848 ATATTATTAATAAATAATACATTTATTAATTAACATTTTATAAACAATAATTCATT 2789
 QY 767 ACCCTTCTGCAGTTCATTAAGCACTTACAGACATTTACTAATAATTAACAAAGAGG 826
 DB 2788 AAATTTATTAATAAATAAACAATTAATTTTTTTTATTTATTAATTAATTTTAAAC 2729
 QY 827 AAGATTTTAACACTTGAAGAGTATGAGATTAAAGCAACATTAAGGGGAGTG 886
 DB 2728 AAATTTATTTTATTAATAAATAAATAATTAATTAATTAATTAATTTATTTAT 2669
 QY 887 TTAATAATTAAGTGTGAACCACTACCTTACTAGTATTAATTAAGAAATTTGAT 946
 DB 2668 TTTTAAATTTTAAATTTTAAAC----ATTTTATTTATTTTATTTTAAATTTAT 2614
 QY 947 CATCACATTAATTAATGCTGATTTAAATTAATGATTAAGTGTATCATTAAGATG 1006
 DB 2613 ATTATTAATTTTATTTTATTTTATTTTATTTTAAATAATTTATTAATTTTAA 2554
 QY 1007 AGAAACCAATAGTCTGCTGCTGATTTTGATTAATTTGTTTCTAGTACTTTCT 1066
 DB 2553 ATTTTATTAATTTTATTAATTAATTAATTTTATTTTATTTTATTTATTAATTTTAA 2494
 QY 1067 CAAGCCTATTAATAAATCTTGATGCTAATGTAATGCTGAAGAAATGCTATGAA 1126
 DB 2493 TATTTTAAACAAATAATTAATAATTAATAATAATAATAATAATAATAATTTTAA 2434
 QY 1127 TTCAATGAATTAATGATTTTCAAAAGTCCAAATCCATCAATGAAATTTAGTACAAA 1186
 DB 2433 AAATTAATTAATTAATTTTATTTTATTTTCAAAAAATAATAATTAATTAATTTAT 2374
 QY 1187 CGTAACCTCAAAATATCTCTATTTTAAATTTTAAAGATTAATAATATCTCTAT 1246
 DB 2373 ATAAAAATTAATAATTAATTAATAATAATAATAATAATAATAATAATAATTAAT 2314
 QY 1247 TTTAAATTTTACAATAATTAATTAACCTGCTCACTTGAATACCAACAATAT 1306
 DB 2313 AAATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2254
 QY 1307 TAATACCTGAATTTTATTTCTTAATTAATTTGAGATCTCTCAATATCTGATTTAT 1366
 DB 2253 TAAAAAATAATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAAT 2194
 QY 1367 TTTAATTTGTGCATATTTCTTATGTTTGAAGTAAACCTTATATCTGCTCAACT 1426
 DB 2193 TTTAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 2134
 QY 1427 AGTAATTC-----ATAATGAGTTTGAAGAGACATGACATCTTGAACATT 1477
 DB 2133 AAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 2074
 QY 1478 GGTTTAACCCTGTGGAATGTTAAGTAATAAATCACTGAGATTAGACCACTAT 1537
 DB 2073 TTTAATAATAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATAAT 2014
 QY 1538 AATATCTCTCTTTGCTTTTAAAAAGTGCATGAAAAATCTCTAGTGAAGCTAGAG 1597
 DB 2013 TAAATTTTATTAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 1954
 QY 1598 TGCTCTGCGCTGTATATCAATTCATTCCAGATGTAAGAACTGCCACTACGAA 1657

DB 1953 TTTTATTTATTAATTAATTAATTAATTTTATTTTATTAATAATTTTATTTATTA 1894
 QY 1658 TAATTTGTCATTAAGACAGTATGTTTAAACAGCTGCCCTTGATGTTTGGCATATAT 1717
 DB 1893 ATAAAAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 1834
 QY 1718 TCGCTCTCTCTCTT- -TTCTCAGCTATTAACAATGACTAATTAAT 1764
 DB 1833 AACAAATTTTATTTATTTATTTTAAATTAATAATAATAATAATAATAATAATAAT 1785

RESULT 7
 AX457067/c 2009 bp DNA linear PAT 06-JUL-2002
 LOCUS Sequence 28 from Patent WO0231186.
 DEFINITION AX457067
 ACCESSION AX457067.1 GI:21715849
 VERSION
 KEYWORDS
 SOURCE synthetic construct
 ORGANISM synthetic construct
 REFERENCE 1
 AUTHORS Berlin, K.
 TITLE Method for the detection of cytosine methylations
 JOURNAL Patent: WO 0231186-A 28 18-Apr-2002;
 EpiGenomics AG (DE)
 FEATURES
 source location/Qualifiers
 1..2009
 /organism="synthetic construct"
 /mol_type="genomic DNA"
 /db_xref="taxon:32630"
 /note="lower strand of DNA sequence 25; after bisulphite treatment and P CR"

BASE COUNT 933 a 0 c 97 g 979 t

Query Match 5.1%; Score 91.4; DB 6; Length 2009;
 Best Local Similarity 47.2%; Pred. No. 0.00015;
 Matches 312; Conservative 0; Mismatches 346; Indels 3; Gaps 1;

QY 782 ATTAAGACACTTTCAGACACTTTTACTAATAATCTAATAAGAGAAATTTTACACT 841
 DB 958 ATATTAAACAACAACATTTTATTTATTTTATTTTAAATTTTAAACAAAAATTAATAATA 899
 QY 842 TAGAGAGTAATGGAGTTAAAGACCAACATTAAGGGGAGTGTAAATTAATGCT 901
 DB 898 ATTTATTTTAAACATTTTAAATTTATTTATTTTAAATTAATAATTAATTAATTAATTTT 839
 QY 902 TGTAAACCACTACCTTTAGTATGATTAAGAAAAATGTAATCATCATTAATTAAT 961
 DB 838 AAATCAAAATTAATTTTAAAAAAATCATTAATAATAATAATAATAATAATAATAATAAT 779
 QY 962 ATTGCTCTTATTTAAATTAATGATAAAGTGTATCATTAAGATTGAGAAAAACAATAGT 1021
 DB 778 AATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAAT 719
 QY 1022 CCGCTCTGATTTTGAATTAATGTTTCTAGTGAATCTTTCTCAAGCCTATTAATA 1081
 DB 718 TTAATTAATTTATTTTAAATTTTAAATTTTAAATAATAATAATAATAATAATAATAAT 659
 QY 1082 ACTTGTATGCTAAATTTGATGCTGGAATAAATGTAATGAAATTCATAGAAATTAAT 1141
 DB 658 ATTTTAAATAAACAATAATTAATTTTAAATTAATTAATTTTATTTTAAATAATAATAATA 599
 QY 1142 GGTATTCAAGTCCAAATTCATCAA--TAGAATTTAGTACAAAACGTAACTCAAAA 1198
 DB 598 TTAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 539
 QY 1199 ATATTCTCTATTTTAAATTTTAAACATTAATAATAATAATAATAATAATAATAATAATA 1258
 DB 538 ATATATATATTTTATTAATTTTAAATAATAATAATAATAATAATAATAATAATAATAATA 479

NOTICE: This sequence may not represent the entire insert of this

clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Ossegeva, K., Moon, P.Y., Zhao, B., Frenken, B., Tien, M., Catane, J.J., and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-362A12; the clone sequenced to the right is RP11-507C12. Actual start of this clone is at base position 1 of RP11-279B11; actual end is at base position 155106 of RP11-279B11.

Data from AC104409 and AC093040 was used to finish this clone, AC104069. Polymorphisms have been identified between AC104409 and AC104069.

The sequence of AC027210 has been incorporated into AC104069.

FEATURES

source Location/Qualifiers

```

1..155106
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="2"
    /map="2"
    /clone="RP11-279B11"
    /clone_lib="RPCI-11"
    858..973
        repeat_region
            /rpt_family="ERV1"
    1100..1124
        repeat_region
            /rpt_family="(TG)n"
    1482..2421
        repeat_region
            /rpt_family="L1"
    2723..33125
        repeat_region
            /rpt_family="L2"
    6121..6181
        repeat_region
            /rpt_family="L1"
    6182..6276
        repeat_region
            /rpt_family="MALR"
    6283..6881
        repeat_region
            /rpt_family="L1"
    6861..6882
        repeat_region
            /rpt_family="AT_rich"
    6935..6971
        repeat_region
            /rpt_family="AT_rich"
    9416..9629
        repeat_region
            /rpt_family="MIR"
    10128..10219
        repeat_region
            /rpt_family="MIR"
    10346..11213
        repeat_region
            /rpt_family="L1"
    10821..10843
        repeat_region
            /rpt_family="CA-rich"

```

```

        /rpt_family="AT_rich"
    11208..12077
        repeat_region
            /rpt_family="L1"
    12114..12173
        repeat_region
            /rpt_family="Alu"
    12176..13052
        repeat_region
            /rpt_family="ERV1"
    13053..13824
        repeat_region
            /rpt_family="ERV1"
    13825..14249
        repeat_region
            /rpt_family="ERV1"
    14232..14785
        repeat_region
            /rpt_family="ERV1"
    14438..14542
        repeat_region
            /rpt_family="CT-rich"
    17665..18541
        repeat_region
            /rpt_family="ERV1"
    18542..18780
        repeat_region
            /rpt_family="Alu"
    18787..18983
        repeat_region
            /rpt_family="L1"
    19002..19280
        repeat_region
            /rpt_family="Alu"
    19654..20155
        repeat_region
            /rpt_family="L1"
    21441..21600
        repeat_region
            /rpt_family="(TA)n"
    21970..21990
        repeat_region
            /rpt_family="AT_rich"
    25585..25609
        repeat_region
            /rpt_family="AT_rich"
    26541..26612
        repeat_region
            /rpt_family="L1"
    26707..26999
        repeat_region
            /rpt_family="Alu"
    26971..26999
        repeat_region
            /rpt_family="(A)n"
    27033..27055
        repeat_region
            /rpt_family="(CA)n"
    27219..27253
        repeat_region
            /rpt_family="(CA)n"
    27428..27813
        repeat_region
            /rpt_family="MALR"
    27883..27922
        repeat_region
            /rpt_family="(TAGA)n"
    27884..28005
        repeat_region
            /rpt_family="L1"
    27978..28007
        repeat_region
            /rpt_family="AT_rich"
    28740..28798
        repeat_region
            /rpt_family="CT-rich"
    30567..30878
        repeat_region
            /rpt_family="Alu"
    30849..30882
        repeat_region
            /rpt_family="(A)n"
    31141..31392
        repeat_region
            /rpt_family="MIR"
    31415..31735
        repeat_region
            /rpt_family="Alu"
    31708..31741
        repeat_region
            /rpt_family="(CAA)n"
    31810..31922
        repeat_region
            /rpt_family="MIR"
    32437..32523
        repeat_region
            /rpt_family="(TATATG)n"
    32463..32531
        repeat_region
            /rpt_family="ERV1"
    32533..32626
        repeat_region
            /rpt_family="MIR"
    32662..32966
        repeat_region
            /rpt_family="Alu"
    32958..33095
        repeat_region
            /rpt_family="CA-rich"

```

| | | | | |
|-----------------------|--------------|--------------------|-----------------|----------------|
| Query Match | 5.1%; | Score 90.8; | DB 9; | Length 155106; |
| Best Local Similarity | 47.2%; | Pred. No. 5.1e-05; | | |
| Matches 308; | Conservative | 0; | Mismatches 342; | Indels 2; |
| | | | | Gaps 1; |

REFERENCE
AUTHORS
 1 (bases 1 to 2500229)
 Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 Carlton J.N., Paine A., Nelson K.E., Bowman S., Paulsen I.T.,
 James K., Eisen J.A., Rutherford K., Salzberg S.L., Craig A.,
 Kres S., Chan M.-S., Nene V., Shallow S.T., Suh B., Peterson J.,
 Agabio L.S., Pertea M., Allen J., Selengut J., Haft D.,
 Mathe M.W., Vaidya A.B., Martin D.M.A., Faltam A.H.,
 Fraunholz M.J., Roos D.S., Ralph S.A., McEdden G.I.,
 Cummings L.M., Subramanian G.W., Mungall C., Venter J.C.,
 Carucci D.J., Hoffman S.L., Newbold C., Davis R.W., Fraser C.M. and
 Barrall B.
TITLE
 Genome sequence of the human malaria parasite *Plasmodium falciparum*
JOURNAL
 Nature 419 (6906), 498-511 (2002)
PUBMED
 12368864
 2 (bases 1 to 2500229)
AUTHORS
 Gardner M.J.
TITLE
 Direct Submission
JOURNAL
 Submitted (13-SEP-2002) The Institute for Genomic Research, 9712
 Medical Center Dr. Rockville, MD 20850, USA
FEATURES
 Location/Qualifiers

60029 L. : 250029

```
/mol_type="genomic DNA"
```

/db xref="taxon:36329"

| Region | 325 | 384 |
|--------|-----|-----|
| ... | ... | ... |

VPC-COMMISSION

```

/rpt_type=candem

```

/locus_tag="PF11_0208"

```

5 locus tag="PF11_0208"

```

```
/]ccys tag="PF1 0208"
```

```
/COORD_SCALE=1
```

```
/protein_id="AAN35792.1"
```

```
/translation="MTTYTTLVLRHGESTWNKENKFTGWTDVPLSEKGEAEIAGK
```

TAKKYGEEOVKIWRSYDIPPKLDKEDNRWPGHNWYKNVPKDALPTECLKDIVER

DEMI KRIZHYVI I DSEFI KYKMEVANOGRAK"

| region | 1404 | .1430 |
|--------|------|-------|
| 1 | 1404 | .1430 |
| 2 | 1404 | .1430 |
| 3 | 1404 | .1430 |
| 4 | 1404 | .1430 |
| 5 | 1404 | .1430 |
| 6 | 1404 | .1430 |
| 7 | 1404 | .1430 |
| 8 | 1404 | .1430 |
| 9 | 1404 | .1430 |
| 10 | 1404 | .1430 |
| 11 | 1404 | .1430 |
| 12 | 1404 | .1430 |
| 13 | 1404 | .1430 |
| 14 | 1404 | .1430 |
| 15 | 1404 | .1430 |
| 16 | 1404 | .1430 |
| 17 | 1404 | .1430 |
| 18 | 1404 | .1430 |
| 19 | 1404 | .1430 |
| 20 | 1404 | .1430 |
| 21 | 1404 | .1430 |
| 22 | 1404 | .1430 |
| 23 | 1404 | .1430 |
| 24 | 1404 | .1430 |
| 25 | 1404 | .1430 |
| 26 | 1404 | .1430 |
| 27 | 1404 | .1430 |
| 28 | 1404 | .1430 |
| 29 | 1404 | .1430 |
| 30 | 1404 | .1430 |
| 31 | 1404 | .1430 |
| 32 | 1404 | .1430 |
| 33 | 1404 | .1430 |
| 34 | 1404 | .1430 |
| 35 | 1404 | .1430 |
| 36 | 1404 | .1430 |
| 37 | 1404 | .1430 |
| 38 | 1404 | .1430 |
| 39 | 1404 | .1430 |
| 40 | 1404 | .1430 |
| 41 | 1404 | .1430 |
| 42 | 1404 | .1430 |
| 43 | 1404 | .1430 |
| 44 | 1404 | .1430 |
| 45 | 1404 | .1430 |
| 46 | 1404 | .1430 |
| 47 | 1404 | .1430 |
| 48 | 1404 | .1430 |
| 49 | 1404 | .1430 |
| 50 | 1404 | .1430 |
| 51 | 1404 | .1430 |
| 52 | 1404 | .1430 |
| 53 | 1404 | .1430 |
| 54 | 1404 | .1430 |
| 55 | 1404 | .1430 |
| 56 | 1404 | .1430 |
| 57 | 1404 | .1430 |
| 58 | 1404 | .1430 |
| 59 | 1404 | .1430 |
| 60 | 1404 | .1430 |
| 61 | 1404 | .1430 |
| 62 | 1404 | .1430 |
| 63 | 1404 | .1430 |
| 64 | 1404 | .1430 |
| 65 | 1404 | .1430 |
| 66 | 1404 | .1430 |
| 67 | 1404 | .1430 |
| 68 | 1404 | .1430 |
| 69 | 1404 | .1430 |
| 70 | 1404 | .1430 |
| 71 | 1404 | .1430 |
| 72 | 1404 | .1430 |
| 73 | 1404 | .1430 |
| 74 | 1404 | .1430 |
| 75 | 1404 | .1430 |
| 76 | 1404 | .1430 |
| 77 | 1404 | .1430 |
| 78 | 1404 | .1430 |
| 79 | 1404 | .1430 |
| 80 | 1404 | .1430 |
| 81 | 1404 | .1430 |
| 82 | 1404 | .1430 |
| 83 | 1404 | .1430 |
| 84 | 1404 | .1430 |
| 85 | 1404 | .1430 |
| 86 | 1404 | .1430 |
| 87 | 1404 | .1430 |
| 88 | 1404 | .1430 |
| 89 | 1404 | .1430 |
| 90 | 1404 | .1430 |
| 91 | 1404 | .1430 |
| 92 | 1404 | .1430 |
| 93 | 1404 | .1430 |
| 94 | 1404 | .1430 |
| 95 | 1404 | .1430 |
| 96 | 1404 | .1430 |
| 97 | 1404 | .1430 |
| 98 | 1404 | .1430 |
| 99 | 1404 | .1430 |
| 100 | 1404 | .1430 |

| | | |
|--------|-------|-------|
| region | 2137. | .2163 |
|--------|-------|-------|

```

region      2291.  .2349

```

2589 2632

752C-CUUM
205A-200A

```
/rpc_type=candem
```

```
/rpt_type=tandem
```

```
/rpt type=tandem
```

```

    _loglik = loglik
    /rnt tyme=tandem

```

| Region | 4056, 4099 |
|--------|------------|
| ... | ... |

<4298. .>4483

<4298. .>4483

4298 - 4483

70000 - 1
70000 - 1
70000 - 1

```
/product="hypothetical protein"
```

```
/db_xref="GI:23496133"
```



```

QY 740 AGTATGGAAGAGAAATGCGATTACCTTTCTGCACTTCAATGAGCACTTACAG 799
Db 71766 ATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 71825
QY 800 CACTTACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 859
Db 71826 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 71865
QY 860 TAAAGCAACACATTAAGGGAGGTGTTAAATGTTGTAACCACTACCTT 919
Db 71866 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 71945
QY 920 TAGTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 979
Db 71946 TATGTCCTA-AATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 72004
QY 980 TATGATTAAGTTGAT-CATTAGATTGAGAAAACCAATGTCCTGCTGATTG 1038
Db 72005 ATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72064
QY 1039 AATTATGTTTCTGATGTTACTTTCTTCAAGCCTATATTAATTAATTAAT 1098
Db 72065 TAAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 72124
QY 1099 TGTATGCTGAAAAAATGCTATGATCAATGAGAAATTAATGATTTCAAGTCAA 1158
Db 72125 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 72184
QY 1159 AATCATCAATGAAATTTAGTCAAAAACGTACTCAAAAATTTCTTATTTAATT 1218
Db 72185 AA-----ATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAAT 72238
QY 1219 TTACACAATTAATTAATTAATTTCTCTTAATTAATTAATTAATTAATTA 1278
Db 72239 AAAAAATGCTCTTAATTAATTAATTAATTTTAATTAATTAATTAATTAAT 72298
QY 1279 GTCACCTTTAGAACCAACCAACATTAATTAATTAATTAATTTCTTAATTT 1338
Db 72299 AGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 72358
QY 1339 GAGATCTGCAATATATCTGATTTTATTTATTTGTCATTTTCTTANGTGTTA 1398
Db 72359 TTTAAATTAATTAATGCTTTTAATTAATTAATTAATTTATTAATTAATTA 72418
QY 1399 GAGTAAACCTTATATCTGTCGCAACAGTAAATCAATATAGAGTTGGAAGAC 1458
Db 72419 TA-TTATTTATTTATTAATTAATTAATTTATTTATTAATTAATTAATTTAT 72477
QY 1459 ATTGACATCTGAAACATTTGTTTAACTTTGTTGGAATGTTAAAGTAAATTA 1518
Db 72478 ATTATATTAATTTATTTATTTATTTATTAATTAATTTATTAATTAATTTT 72537
QY 1519 AGAATTATGACCATTAATTAATTAATTAATTTCTGTTTAA 1561
Db 72538 ATTATTTATTAATTAATTAATTAATTTATTTATTTATTTA 72580

RESULT 11
LOCUS CNG01RGG 176368 bp DNA linear PRI 13-NOV-2001
DEFINITION Human chromosome 14 DNA sequence BAC R-93311 of library RPI-11
from chromosome 14 of Homo sapiens (human), complete sequence.
ACCESSION AL157971.4 GI:16944346
VERSION AL157971.4
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 176368)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brettier,P., Catolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,

```

TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL

Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C., Gyapay, G., Saurin, W. and Weissenbach, J. Sequencing of the human chromosome 14 Unpublished 2 (bases 1 to 176368) Genoscope. Direct Submission Submitted (08-NOV-2001) Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - FRANCE (E-mail : segref@genoscope.cns.fr - Web : www.genoscope.cns.fr) On Nov 15, 2001 this sequence version replaced gi:14715170. ----- Genome Center Center: Genoscope / Centre National de Sequencage Center code: GS Web site: http://www.genoscope.cns.fr/ Contact: segref@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
----- Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 7.53x in Q20 bases, sum-of-contigs

Overall quality chart :
Range : bases

| | |
|---------|-------|
| 0 | 5 |
| 1 - 9 | 25 |
| 10 - 19 | 67 |
| 20 - 29 | 185 |
| 30 - 39 | 2335 |
| 40 - 49 | 5835 |
| 50 - 59 | 7981 |
| 60 - 69 | 20287 |
| 70 - 79 | 49616 |
| 80 - 89 | 89832 |
| 90 - 99 | |

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES

source

1..176368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone_id="R-93311"
/clone_11b="RPI-11"

BASE COUNT 49878 a 31876 c 34551 g 60063 t
ORIGIN

Query Match 5.1%; Score 90.2; DB 9; Length 176368;
Best Local Similarity 47.7%; Pred. No. 6e-05;
Matches 326; Conservative 0; Mismatches 353; Indels 4; Gaps 2;

```

QY 885 TGTAAATTAATTAATGTTGTTAACAACCACTTACCTTACGATTAATTAAGAAATTTGTA 944
Db 56913 TATTATATTAATTAATGTTAATTAATTTTATTAATTAATTAATTAATTAATTAAT 56854
QY 945 ATCATCACTTAATTAATTAATGCTTATTAATTAATTAATTAATTAATTAATTAATTA 1004
Db 56853 ATTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 56794
QY 1005 TGAGAAAACCAATATGCTGCTGCTGATTTTGAATTTTGAATTTTCTTCTAGTTACTTTC 1064
Db 56793 ATTACATTTTACATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 56734
QY 1065 TTCAGACCTTAATTAATTAATTT--TGTATGCTAAATGTTATGCTGGAAAAATGCTGA 1121
Db 56733 TATATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 56674
QY 1122 ATGAATTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1181
Db 56673 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 56614

```

QY 1182 CAAACGTAACCTCAAAATATCTCTATTTTAAATTTTACACAAATATTAATAATCTT 1241
DB 56613 TAT 56555
QY 1242 CTTATTTTAAATTTTACAAAT 1301
DB 56554 ATTAT 56495
QY 1302 AATTTTAACTTATGATATTTTATCTTAATATATTTTGAATCTTCAATATATCTGATA 1361
DB 56494 AAT 56435
QY 1362 TTTATTTTATATTTTGTGCATATTTTCTTATNGTTTGAATTTTACCTTTATCTTGTC 1421
DB 56434 TTTAT 56375
QY 1422 AAATCTATATATCAATATATATGAGTTTGTGAAGACACATTTGACATCTTGAACATTTGTT 1481
DB 56374 AAT 56315
QY 1482 TTTACCTTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 1541
DB 56314 GAT 56255
QY 1542 TACTTCTTTTGTCTTTTAAAAA 1564
DB 56254 TATATGTACACCTCTTGCAATA 56232

RESULT 12
AC131649 181976 bp DNA linear PRI 25-AUG-2002
LOCUS Homo sapiens chromosome 16 clone RP11-609N14, complete sequence.
DEFINITION AC131649 AC016139
AC131649.1 GI:22475305
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 181976)
REFERENCE DOB Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 181976)
DOE Joint Genome Institute.
REFERENCE Direct Submission
JOURNAL Submitted (25-AUG-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Aug 25, 2002 this sequence version replaced GI:20279329.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
Source
1. 181976
/organism="Homo sapiens"

BASE COUNT 51909 a 37229 c 38430 g 54408 t
ORIGIN
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-609N14"

Query Match 5.1%; Score 90.2; DB 9; Length 181976;
Best Local Similarity 48.8%; Pred. No. 6e-05;
Matches 306; Conservative 0; Mismatches 313; Indels 8; Gaps 2;

QY 912 ACTACCTTATAGATATATATAGAAATTTGATATATATATATATATATATATATATATAT 971
DB 174813 ATTAT 174872
QY 972 TTTAAAT 1031
DB 174873 TAT 174932
QY 1032 ATTTTGAATTTATGTTTCTTATGTTTCTTCTTCAAGCTTATATATATATATATATAT 1091
DB 174933 TAT 174992
QY 1092 GCTAATATGATGCTGGAAGAAATATGATATATATATATATATATATATATATATATAT 1151
DB 174993 TAT 175052
QY 1152 AGTCCAAATTCATCATATGAAATTTAGTACAAAAGTACCTCAAAATATCTCTTAT 1211
DB 175053 AAT 175112
QY 1212 TTAATTTTACACAAAT 1271
DB 175113 AT 175172
QY 1272 ATACCTGTCACCTTTAGAAATACCAACAATATATATATATATATATATATATATATAT 1331
DB 175173 ATAA-----AT 175228
QY 1332 TTAATTTTGAATCTCTCAAT 1391
DB 175229 TAAAT 175284
QY 1392 TGTTTTAAAGTTTAACTTATATCTGTGCAACAGTAAATTCATATATATATATATATAT 1451
DB 175285 TAT 175344
QY 1452 AGGACACATTTGACATCTTGAACAATGTTTAACTTGTGGAATGTTTAAAGTAAATTA 1511
DB 175345 TAT 175404
QY 1512 AACATTCAGATTTTATGACCATCTATTA 1558
DB 175405 AAT 175431

RESULT 13
AC117444/c 104853 bp DNA linear PRI 31-JUN-2002
LOCUS Homo sapiens 3 BAC RP11-215L17 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
DEFINITION AC117444
AC117444.6 GI:22024337
VERSION
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 104853)
REFERENCE Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,

Buhay, C., Burch, P., Burkett, C., Butrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Delaney, K.R., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Din, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.M., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotte, M., Falla, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, M., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, U., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Holy, J., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Jouda, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korval, J., Kovar, C., Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, J.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louiegeed, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, N.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, B., Nwokenkwo, S., Oghu, M., Okumura, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojoubokan, I., Rolfe, N., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I., Sodergren, E., Sonakke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svalok, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, D., Zorrilla, S., Zyltor, S.L., Weinstein, G., and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 104853)
Worley, K.C.
Direct Submission
Submitted (10-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 104853)
Worley, K.C.
Direct Submission
Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 104853)
Worley, K.C.
Direct Submission
Submitted (31-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 31, 2002 this sequence version replaced gi:21908338.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green,

unpublished) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu/quality.info/genbank.annotation.htm>.

| FEATURES | source |
|---------------|---|
| misc_feature | Location/Qualifiers |
| repeat_region | 1..104853 |
| repeat_region | /organism="Homo sapiens" |
| repeat_region | /mol_type="genomic DNA" |
| repeat_region | /db_xref="taxon:9606" |
| repeat_region | /chromosome="3" |
| repeat_region | /clone="RP11-215L17" |
| repeat_region | 1..2000 |
| repeat_region | /note="overlaps bases 154275..156277 of clone AC107015" |
| repeat_region | /function="clone overlap" |
| repeat_region | 246..266 |
| repeat_region | /rpt_family="A)n" |
| repeat_region | 411..954 |
| repeat_region | /rpt_family="MLT2A" |
| repeat_region | 1576..1598 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | complement(12562..3082) |
| repeat_region | /rpt_family="MLT1G" |
| repeat_region | 4553..4922 |
| repeat_region | /rpt_family="MLT1A1" |
| repeat_region | 5010..5035 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 5079..5406 |
| repeat_region | /rpt_family="L2" |
| repeat_region | 6323..6351 |
| repeat_region | /rpt_family="A)n" |
| repeat_region | 6948..7099 |
| repeat_region | /rpt_family="MIR" |
| repeat_region | 7338..7447 |
| repeat_region | /rpt_family="MER94" |
| repeat_region | 8267..8294 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 9151..9242 |
| repeat_region | /rpt_family="MLT1U" |
| repeat_region | 9296..9509 |
| repeat_region | /rpt_family="MLT1U2" |
| repeat_region | 10028..10091 |
| repeat_region | /rpt_family="(CATATA)n" |
| repeat_region | complement(13636..14644) |
| repeat_region | /rpt_family="MER45B" |
| repeat_region | 15307..15335 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 15593..15628 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | 15747..16002 |
| repeat_region | /rpt_family="A1uy" |
| repeat_region | 16122..16353 |
| repeat_region | /rpt_family="MIR" |
| repeat_region | 17470..17502 |
| repeat_region | /rpt_family="AT_rich" |
| repeat_region | complement(18294..18632) |


```

repeat_region      /rpt_family="L1M4"
19255..19775
/rpt_family="L1PA13"
repeat_region      20757..20815
/rpt_family="CT-rich"
repeat_region      complement(12111)..22252
/rpt_family="L1PA5"
22521..22798
STS
repeat_region      /standard_name="D11S3206"
27253..27294
/rpt_family="AT-rich"
repeat_region      complement(12752)..28145
/rpt_family="L1MC1"
28155..28203
/rpt_family="AT-rich"
repeat_region      complement(128585)..28682
/rpt_family="U6"
29288..29952
/rpt_family="MER76"
30941..30960
/rpt_family="(TTTC)n"
repeat_region      complement(30963)..31250
/rpt_family="A15y"
repeat_region      complement(31825)..31911
/rpt_family="i2"
repeat_region      complement(32569)..32880
/rpt_family="A1uS4"
repeat_region      33002..33226
/rpt_family="LTR59"
repeat_region      33247..33731
/rpt_family="MER72"
repeat_region      complement(33796)..34214
/rpt_family="i2"
repeat_region      complement(34233)..35306
/rpt_family="MER11C"
35706..35894
/rpt_family="MR"
repeat_region      36206..36244
/rpt_family="AT-rich"

```

| Query Match | 5.0% | Score 90; | DB 95; | Length 104853; |
|-----------------------|---|--------------------|-----------|----------------|
| Best Local Similarity | 45.4% | Pred. No. 7.5e-05; | | |
| Matches 399; | Conservative 0; | Mismatches 475; | Indels 4; | Gaps 2; |
| QY 888 | TAAATAATATGCTGGTTGAACCAACCACTACCTTAGTAAGATTATTAAGAAATTTGTAATC | 947 | | |
| Db 40503 | TATATAAATATATATATATATATATATATATAAATAATATATATATATATATATATATATAAATAATAT | 40444 | | |
| QY 948 | ATCATATATATATATATGTCCTTTTAAATTTTGAATTAAGTTGATTCATTTAAGTTGA | 1007 | | |
| Db 40443 | TATATATATATATATATAAATAATATATATATATATATATATATATATATAAATAATATATATATAT | 40384 | | |
| QY 1008 | GAATAACCAATATAGTCCTGCTCTGATTTTGGAAATTTGTTTCATAGTACTTTCTTC | 1067 | | |
| Db 40383 | ATATATATAAATATATATATATATATATATATATATAAATAATATATATATATATATATAAATAA | 40322 | | |
| QY 1068 | AAGCTTATATATAAATCTTTGAATGCTGAATTTGATGCTGGAAAAATTTGTATATGAAT | 1127 | | |
| Db 40323 | TATATATATATATATATATATATATAAATAATATATATATATATATATATATAAATAATATATATAT | 40264 | | |
| QY 1128 | TCAATAGAAATATATGATTTTCAAGAGCCAAATCCATCAATGAAATTTAGTAGCAAAAC | 1187 | | |
| Db 40263 | TATATATATAAT | 40206 | | |
| QY 1188 | GTAATCTAAAAATATCTCTTATATTTAAATTTTCAACAATATATATAAATAATCTCTTAT | 1247 | | |
| Db 40203 | TTATATATATATATATATATATATACATTTTATATATATATATATATATATATATATATATATATA | 40144 | | |
| QY 1248 | TTAAATTTTCAATATATATATATATATATATATACCGTCACCTTAGAGATCCACCAACATAT | 1306 | | |
| Db 40143 | TATATAATATATATAAATAATCTATATAAATAATATATATATATATATATAAATAATATATATATATAA | 40084 | | |
| QY 1307 | TAAATCTTAGATATTTTATTTCTTAAATAATTTTGAGATCTCTCAATATATCTGATATTTAT | 1366 | | |

Dd 40083 TAAATATATACTATATATATTATTATTAATAATGTATACACTTATATATAAATATATAT 40024

Qy 1367 TTATATTTGGTCAGT--TTCTCATGTTTTAGAGTAAACCCTTATCTTGCTCA 1423

Dd 40023 TTATGATCTMAAATTATATGTAATATATAAAAAATTATTTATCTATAATTTAT 39966

Qy 1424 ACTAGTAATCAATATATGAGTTTGTGAAGACACATGCATCTTGAAACATTTGTTT 1483

Dd 39963 ATATAAATATAAAAATATGATTTTATTTATATAGAATAAATATATATTAAGTATAT 39904

Qy 1484 AACTTGTGGATGTTAAAGGTAATPAAAACATTCGAATTAGCCATCTATTAATATA 1543

Dd 39903 ATATCATTTATTAATAAACAATATGATTCATTATAATATAATATTATTAATATT 39844

Qy 1544 CTCCTCTGCTTTTAAAAAGGTGATGAAATGCTATGAGTAGAGTGCTCT 1603

Dd 39843 ATATATTTATTTAGTATATAATTTATATATTTATTTATATTTGGAATATATATTTTAT 39786

Qy 1604 GCTGCCCTGTATATCAATTCATTCATTCAGATGTAGAACTGCCACRACGATATATA 1663

Dd 39783 TATAAATATATTTTATTTATTTATATACAAAACCTGATATTTTATTTATATGAGAAATA 39724

Qy 1664 GTCAATTAAGCAGTATGTAACACGCTCCCTGCATGTTTTTGGCATATATTCGCT 1723

Dd 39723 TATATATTAATATATTTATTTATATATTTATTTATATAATATTAATATGTTTTGTTA 39666

Qy 1724 TCTTTCTTTTCTCAAGTATAAACACATCACTATT 1761

Dd 39663 TGATATAATATATAACATCACTTTCTCGATTTTAT 39626

```

RESULT 14
BX248102/c
LOCUS BX248102.1
DEFINITION Danio rerio clone CH211-243f3, *** SEQUENCING IN PROGRESS ***, 2
ACCESSION BX248102
VERSION BX248102.5 GI:29538884
KEYWORDS HTGS PHASE1, HTGS_DRAFT, HTGS_FULLTOP.
SOURCE Danio rerio (zebrafish)
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 150336)
McLaren,S.
Direct Submission
Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
zfish-help@sanger.ac.uk Clone request: clonerequest@sanger.ac.uk
On Apr 3, 2003 this sequence version replaced gi:28301010.

COMMENT
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: ZC24333
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 150068 bases at least Q40
Consensus quality: 150124 bases at least Q30
Consensus quality: 150124 bases at least Q20
Insert size: 150136; sum-of-contigs
Insert size: 158004; 3.9% error; agarose-fp
Quality coverage: 8.79x in Q20 bases; sum-of-contigs Quality
coverage: 8.35x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 130320: contig of 130320 bp in length
 * 130321 130420: gap of 100 bp
 * 130421 150236: contig of 19816 bp in length.

FEATURES

source

1. 150236

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="CH211-243F3"

/clone_11b="CHORI-211"

1. 130320

/note="assembly_fragment:01135"

fragment_chain:1

clone_end:SP6

vector_side:left"

130421..150236

fragment_chain:1

/note="assembly_fragment:00144"

clone_end:T7

vector_side:right"

BASE COUNT 47331 a 26020 c 27386 g 49399 t 100 others

ORIGIN

Query Match

Best Local Similarity 47.0%; Pired. No. 6.8e-05;
 Matches 314; Conservative 0; Mismatches 350; Indels 4; Gaps 1;

786 GAGCACTTACAGACCTTTTACTTAAATCTACAAAGAGAGATTTTAAACACTTGA 845

131133 GTGCAATATGCGCTTTTAAAGATTAATTAACAACTGTATCTTAATATAGCTGT 131074

846 GAAGTATGGAGTTAAGACCAACATTAAGGGAGCTTAAATTAATGTTGTA 905

131073 ATTATGCCACCAAGTTATAGTACTTAAACCTGCTTACCAACTTTTGA 131014

906 ACCACCACTACCTTTAGTATGATTTATAGAAAATGTTATCATCATTAATTTG 965

131013 GCACCACTTACTAGCTTATATATATATATATATATATATATATATATAT 130954

966 TCCTTATTTAAATTAATGATTAAGCTGTATCATTTAGATTGAGAAAACCAATAGCC 1025

130953 ATATTAT 130894

1026 GCTTGAATTTTGAATTTGTTTCTATGTTACTTTCTCAAGCTATATAAACTT 1085

130893 TTATTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTT 130834

1086 TGTAACTTAATTTGATGCTGGAAGAAAATGTTAATGAATTCATAGAAATTA 1145

130833 TAT 130774

1146 TTTCAAGTCCAAATTCATCAATAGAAATTTAGTACAAAACGTAATCAAAATTTCT 1205

130773 TATTAATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 130714

1206 CTATTTTAAATTTTACAAATATA---AAAAATCTCTTATTTTAAATTTTCAAT 1261

130713 ATTAAT 130654

1262 AATTAATTTTACCTGCTCCTTTAGAAATACCAACAACAAATTAATCTAGAAAT 1321

130653 ATTAAT 130594

1322 TTATCTTAAATTTTGAATCTCTCAATATATGATATTTATTTATTTTGTCTCA 1381

130593 TTAT 130534

1382 TATTTCTTATGTTTAAAGTTAAACCTTATATCTTGCTCAACTAGTAATCAATATAT 1441

Db 130533 TTATTTAT 130474

Qy 1442 GAGTTTCT 1449

Db 130473 TATATATAT 130466

RESULT 15

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CENTER

CONTACT

PROJECT

SUMMARY

STATISTICS

ASSEMBLY

PROGRAM

DYE

TERMINATOR

CONSISTENCY

QUALITY

INSERT

SIZE

SUM-OF-CONTIGS

INSERT

SIZE

SUM-OF-CONTIGS

INSERT

SIZE

SUM-OF-CONTIGS

INSERT

SIZE

SUM-OF-CONTIGS

INSERT

SIZE

SUM-OF-CONTIGS

INSERT

SIZE

SUM-OF-CONTIGS

FEATURES

source

1. 153326

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="CH211-240A8"

/clone_11b="CHORI-211"

1. 19915

/note="assembly_fragment:01135"

clone_end:T7

NOTE: This is a 'working draft' sequence. It currently

consists of 5 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

19915: contig of 19915 bp in length

19916 20015: gap of 100 bp

20016 25492: contig of 5477 bp in length

25493 25592: gap of 100 bp

25593 36617: contig of 11025 bp in length

36618 36717: gap of 100 bp

36718 135889: contig of 99272 bp in length

135890 136089: gap of 100 bp

136090 153326: contig of 17237 bp in length.

```
misc_feature      vector size:left  
20016.._25492  
/note="assembly_fragment:00605  
fragment_chain:1"  
misc_feature      23593.._36617  
/note="assembly_fragment:01486  
fragment_chain:1"  
misc_feature      36718.._135989  
/note="assembly_fragment:01039  
fragment_chain:1"  
misc_feature      136090.._153326  
/note="assembly_fragment:01628  
fragment_chain:1  
clone_end:SP6  
vector size:right"  
BASE COUNT      50335 a 27839 c 26492 g 48257 t 403 others  
ORIGIN
```

```
Query Match      5.0%; Score 90; DB 2; Length 153326;  
Best Local Similarity 47.0%; Pred. No. 6,7e-05;  
Matches 314; Conservative 0; Mismatches 350; Indels 4; Gaps 1;
```

```
QY 786 GAGCACTTACAGACCTTTTACCAAACTACCAAGAGAGATTTTACCACTTGA 845  
DB 19104 GTGCACATATGCGATTTTAAAGTAAATAAACAACCTGTACTTAATATAGGCTGT 19163  
QY 846 GAAATTAATGGAGTTAAAGCAACATTAAGGGAGGTAAATATGTGTGA 905  
DB 19164 ATTATGCCACAGTTAATAGTACCTGATTAACCTGAGTACCAACTTTTGA 19223  
QY 906 ACCACCACTACCTTTAGTATATTAAGAAATGTATCATCATTAATATTTG 965  
DB 19224 GCAGCCATTAAGTCTATTAATTAATTAATTAATTAATTAATTAATTAAT 19283  
QY 966 TCCATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19283  
DB 19284 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19343  
QY 1026 GTCTGATTTTGAATTAATGTTTCTAGTTACTTTCTCAAGCCATTAATAAACT 1085  
DB 19344 TTATTTTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19403  
QY 1086 TGTATGCTAAATGTGCTGAAAAAATGTATATGAATCAATAGAAATTAATGTA 1145  
DB 19404 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19463  
QY 1146 TTCAAGTCCAAATCCATCAATAGAAATTAATTAATTAATTAATTAATTAAT 1205  
DB 19464 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19523  
QY 1206 CTAATTTTAATTTTACAAATATA---AAATATCTCTATTTTAATTTTACAT 1261  
DB 19524 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19583  
QY 1262 AATATAATTTACCTGCTGACCTTAGAATACCAACAAATTAATTAATTAATTAAT 1321  
DB 19584 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19643  
QY 1322 TTATCTTAATTAATTTGAGATCTGCAATATGATATTTATTTATATTTGCTCA 1381  
DB 19644 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19703  
QY 1382 TATTTCTTAATTTGAGTTAGATTAACCTTAATCTGCTCAACTAGTAATTAATTAAT 1441  
DB 19704 TTATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19763  
QY 1442 GAGTTTGT 1449  
DB 19764 TATATTAT 19771
```

This Page Blank (uspto)

Wed Nov 26 09:11:05 2003

us-09-831-083-1.fst

Page 1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 16:23:54 : Search time: 2616 Seconds
(Without alignments)
16565.321 Million cell updates/sec

Title: US-09-831-083-1
Perfect score: 1783
Sequence: 1 atccacctctgacctctga.....tagagcgcacacgctgacc 1783

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 segs, 1215238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST.*
1: em_estda:*
2: em_estlum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hnc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlum:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_iny:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rnd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 24 | 1.3 | 625 | 28 | AZ978994 |
| 2 | 23 | 1.3 | 189 | 10 | BA422474 |
| 3 | 23 | 1.3 | 575 | 12 | BU562777 |
| 4 | 23 | 1.3 | 600 | 12 | BU563982 |

| | | | | | | | | |
|---|----|----|-----|------|----|-----------|-----------|------------|
| C | 5 | 23 | 1.3 | 708 | 29 | BZ323523 | BZ323523 | 1a79c09.g |
| C | 6 | 22 | 1.2 | 342 | 28 | BH780793 | BH780793 | l2mb020f0 |
| C | 7 | 22 | 1.2 | 429 | 14 | CA571794 | CA571794 | K0524H02- |
| C | 8 | 22 | 1.2 | 479 | 13 | BQ562120 | BQ562120 | H4074C02- |
| C | 9 | 22 | 1.2 | 495 | 28 | BH787917 | BH787917 | f2mb020f0 |
| C | 10 | 22 | 1.2 | 592 | 10 | BF198774 | BF198774 | 248732.MA |
| C | 11 | 22 | 1.2 | 653 | 29 | BZ421643 | BZ421643 | b232f07.b |
| C | 12 | 22 | 1.2 | 661 | 14 | CA995790 | CA995790 | rg06a11.y |
| C | 13 | 22 | 1.2 | 1136 | 29 | CC283036 | CC283036 | CH261-112 |
| C | 14 | 21 | 1.2 | 196 | 9 | AU060103 | AU060103 | AU060103 |
| C | 15 | 21 | 1.2 | 272 | 13 | BQ851931 | BQ851931 | QGB16K21. |
| C | 16 | 21 | 1.2 | 371 | 28 | A2845955 | A2845955 | 2M0145P22 |
| C | 17 | 21 | 1.2 | 423 | 10 | BG008892 | BG008892 | CM2-GN022 |
| C | 18 | 21 | 1.2 | 438 | 28 | AZ003018 | AZ003018 | RPCI-23-3 |
| C | 19 | 21 | 1.2 | 488 | 9 | A1700637 | A1700637 | we8908.x |
| C | 20 | 21 | 1.2 | 501 | 12 | B1867863 | B1867863 | fl64g09.y |
| C | 21 | 21 | 1.2 | 510 | 10 | BF814093 | BF814093 | RC3-C1019 |
| C | 22 | 21 | 1.2 | 511 | 12 | BM336511 | BM336511 | MEST195-A |
| C | 23 | 21 | 1.2 | 531 | 14 | CB162751 | CB162751 | K-BST0223 |
| C | 24 | 21 | 1.2 | 534 | 9 | AL922682 | AL922682 | AL922682 |
| C | 25 | 21 | 1.2 | 542 | 29 | BZ781272 | BZ781272 | 1126a03.g |
| C | 26 | 21 | 1.2 | 557 | 13 | BQ851931 | BQ851931 | QGB16A17. |
| C | 27 | 21 | 1.2 | 562 | 29 | BZ779081 | BZ779081 | 1186e05.b |
| C | 28 | 21 | 1.2 | 603 | 10 | BF507982 | BF507982 | UI-H-B14- |
| C | 29 | 21 | 1.2 | 622 | 13 | BQ850954 | BQ850954 | QGB14A08. |
| C | 30 | 21 | 1.2 | 636 | 29 | BX168297 | BX168297 | Danilo rer |
| C | 31 | 21 | 1.2 | 647 | 12 | B37371907 | B37371907 | B37371907 |
| C | 32 | 21 | 1.2 | 648 | 12 | B3430558 | B3430558 | B3430558 |
| C | 33 | 21 | 1.2 | 654 | 29 | BZ869284 | BZ869284 | CH240-.248 |
| C | 34 | 21 | 1.2 | 662 | 12 | BG776697 | BG776697 | 60263858 |
| C | 35 | 21 | 1.2 | 671 | 28 | A2516262 | A2516262 | RPCI-11-3 |
| C | 36 | 21 | 1.2 | 674 | 14 | CA448878 | CA448878 | UI-H-E10- |
| C | 37 | 21 | 1.2 | 694 | 13 | BU028797 | BU028797 | QHR4f11.y |
| C | 38 | 21 | 1.2 | 699 | 9 | AU060686 | AU060686 | AU060686 |
| C | 39 | 21 | 1.2 | 703 | 29 | AG142042 | AG142042 | Pan treg1 |
| C | 40 | 21 | 1.2 | 708 | 13 | BU719850 | BU719850 | SLM2AHF01 |
| C | 41 | 21 | 1.2 | 713 | 12 | B3773732 | B3773732 | B3773732 |
| C | 42 | 21 | 1.2 | 715 | 12 | B3386295 | B3386295 | B3386295 |
| C | 43 | 21 | 1.2 | 728 | 13 | BQ988528 | BQ988528 | QGR15C17. |
| C | 44 | 21 | 1.2 | 730 | 13 | BQ995401 | BQ995401 | QGR5C009.y |
| C | 45 | 21 | 1.2 | 731 | 12 | BT402629 | BT402629 | BT402629 |

ALIGNMENTS

RESULT 1
LOCUS AZ978994
DEFINITION 2M0255104R Mouse 10kb plasmid UUGC2M library Mus musculus genomic
clone UUGC2M0255104.R, genomic survey sequence.
ACCESSION AZ978994
VERSION AZ978994.1 GI:13850221
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE
1 (bases 1 to 625)
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmood,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A.
and Wright,D., Weiss,R.

TITLE
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL
COMMENT
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asterids; Lamiales; Solanales; Convolvulaceae; Ipomoea.

1 (bases 1 to 575)

Hoshino,A., Seki,M., Shin-I,T., Carninci,P., Kamiya,A., Shiraki,T., Nitaaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.

ESTs of Japanese morning glory

Unpublished

Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@gens.nig.ac.jp.

Location/Qualifiers

1. 575

/organism="Ipomoea nil"

/mol_type="mRNA"

/cultivar="Tokyo-kokei standard"

/db_xref="taxon:35883"

/clone="j1m37p16"

/issue_type="mixture of flower and flower bud"

/clone_lib="Ipomoea nil mixture of flower and flower bud"

141 a 146 c 93 g 195 t

BASE COUNT

141 a 146 c 93 g 195 t

ORIGIN

Query Match 1.3%; Score 23; DB 12; Length 575;

Best Local Similarity 100.0%; Pred. No. 8.3;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1718 TCCGCTCTCTTTCTTTCTTCAC 1740

|||||

527 TCCGCTCTCTTTCTTTCTTCAC 549

|||||

RESULT 4

LOCUS BJS63982 600 bp mRNA linear EST 18-DEC-2002

DEFINITION cDNA clone jm40n16 5', mRNA sequence.

ACCESSION BJS63982

VERSION BJS63982.1 GI:27245802

KEYWORDS EST.

SOURCE

ORGANISM

Ipomoea nil (Japanese morning glory)

Ipomoea nil

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asterids; Lamiales; Solanales; Convolvulaceae; Ipomoea.

1 (bases 1 to 600)

Hoshino,A., Seki,M., Shin-I,T., Carninci,P., Kamiya,A., Shiraki,T., Nitaaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.

ESTs of Japanese morning glory

Unpublished

Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@gens.nig.ac.jp.

Location/Qualifiers

1. 600

/organism="Ipomoea nil"

/mol_type="mRNA"

/cultivar="Tokyo-kokei standard"

/db_xref="taxon:35883"

/clone="j1m40n16"

/issue_type="mixture of flower and flower bud"

/clone_lib="Ipomoea nil mixture of flower and flower bud"

150 a 154 c 96 g 200 t

BASE COUNT

150 a 154 c 96 g 200 t

ORIGIN

Query Match 1.3%; Score 23; DB 12; Length 600;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1718 TCCGCTCTCTTTCTTTCTTCAC 1740

|||||

527 TCCGCTCTCTTTCTTTCTTCAC 549

|||||

RESULT 5

LOCUS BZ233523/c 708 bp DNA linear GSS 06-NOV-2002

DEFINITION ia79c09.g1 WGS-ZmaySF (JM107 adapted methyl filtered) Zea mays

genomic clone ia79c09 5', genomic survey sequence.

ACCESSION BZ233523

VERSION BZ233523.1 GI:24702108

KEYWORDS GSS.

SOURCE

ORGANISM

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 708)

Rabinowitz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zutavern,T., McCombie,W.R. and Martienssen,R.A.

Genomic shotgun sequences from Zea mays (methyl-filtered)

Unpublished

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: ia79 row: c column: 09

Seq primer: -21M3UnivRev

Class: shotgun

High quality sequence stop: 708.

Location/Qualifiers

1. 708

/organism="Zea mays"

/mol_type="genomic DNA"

/cultivar="B73"

/db_xref="taxon:4577"

/clone="ia79c09"

/lab_host="JM107 or DH5a"

/clone_lib="WGS-ZmaySF (JM107 adapted methyl filtered)"

/note="Organ: Immature ears; Site_1: Xba I; Site_2: Xba I; The vector was digested with Xba I and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephader. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, b/g reads in pUC19). The same ligation was transformed in either JM107 or DH5a."

172 a 178 c 139 g 219 t

BASE COUNT

172 a 178 c 139 g 219 t

ORIGIN

Query Match 1.3%; Score 23; DB 29; Length 708;

Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1211 TTAAATTTTACACATATATA 1233

|||||

506 TTAAATTTTACACATATATA 484

|||||

RESULT 6

LOCUS BH780793 342 bp DNA linear GSS 28-MAR-2002

DEFINITION f2mb020f010g10f0 f2mb filtered library Zea mays genomic clone f2mb020f010g10 5', genomic survey sequence.

ACCESSION BH780793
 VERSION BH780793.1 GI:19783834
 KEYWORDS GSS.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 342)
 AUTHOR Budiman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.
 TITLE Geneshresher methylation filtered genomic sequences from maize
 JOURNAL Unpublished
 COMMENT Contact: Bedell JA
 Orion Genomics, LLC
 4041 Forest Park Ave, St. Louis, MO 63108, USA
 Tel: 314 615 6979
 Fax: 314 615 5975
 Email: jbedell@oriongenomics.com
 Plate: fmb020f010 row: g column: 10
 Seq primer: M13 forward
 Class: shotgun
 High quality sequence stop: 342.
 FEATURES
 source
 1..342
 /organism="Zea mays"
 /mol_type="Genomic DNA"
 /cultiVar="MO17"
 /db_xref="taxon:4577"
 /clone_lib="fmb020f010g10"
 /note="Organ: leaf; Vector: pGCSK(-); Site: 1; HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pGCSK(-) vector and electroporated into E. coli cells."
 BASE COUNT 111 a 59 c 69 g 103 t
 ORIGIN
 Query Match 1.2%; Score 22; DB 28; Length 342;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 750 GAAGGAATGTCAGTACTT 771
 118 GAAGGAATGTCAGTACTT 139
 RESULT 7
 CAS71794/c 429 bp mRNA linear EST 19-NOV-2002
 LOCUS K0524H02-SN NIA Mouse Hematopoietic Stem Cell (lin-/c-Klt-/Sca-1+)
 DEFINITION CDNA Library (Long) Mus musculus CDNA clone NIA:K0524H02
 IMAGE:30066229 5', mRNA sequence.
 ACCESSION CAS71794
 VERSION CAS71794.1 GI:25116495
 KEYWORDS EST
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 429)
 AUTHOR Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G., Taub, D., Longo, D.L., Keller, J. and Ko, M.S.H.
 TITLE Systematic Analyses of NIA Mouse Hematopoietic Stem Cell (Lin-/c-Klt+/Sca-1+) CDNA Library (Long)
 JOURNAL Unpublished
 Other_ESTs: K0524H02-3
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.grc.nia.nih.gov
 Plate: K0524 row: H column: 02

Seq primer: M13 Reverse
 High quality sequence stop: 429
 POLYA=No.
 FEATURES
 source
 location/Qualifiers
 1..429
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="G57BL/6NCr"
 /db_xref="taxon:10090"
 /clone="NIA:K0524H02-5N"
 /tissue_type="Hematopoietic Stem Cell (lin-/c-Klt+/Sca-1+)"
 /dev_stage="Age approx. 10 weeks old"
 /lab_host="DH10B"
 /clone_lib="NIA Mouse Hematopoietic Stem Cell (lin-/c-Klt+/Sca-1+) CDNA Library (Long)"
 /note="Vector: pSPORT1 (Invitrogen); Site: 1; SalI; Site: 2; NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). PMID: 11544199). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dt) primer (Invitrogen): 5'-pGACTAGTCTAGATCGGAGCGCGCCCTTTTCTTTTCTTTT-3' from 4.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lops-linker II-SalI, purified by phenol/chloroform, and separated from free linkers by centrifugation. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and centrifugation. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.7 kb. The library was constructed by Yulan Piao (NIA)."
 BASE COUNT 99 a 102 c 88 g 140 t
 ORIGIN
 Query Match 1.2%; Score 22; DB 14; Length 429;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1560 AAAAAGTGTGATGAAATGCT 1561
 61 AAAAAGTGTGATGAAATGCT 40
 RESULT 8
 B0562120/c 479 bp mRNA linear EST 20-JUN-2002
 LOCUS H4074C02-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone
 DEFINITION H4074C02 5', mRNA sequence.
 ACCESSION B0562120
 VERSION B0562120.1 GI:21463006
 KEYWORDS EST
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 479)
 AUTHOR VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin, P.R., Stagg, C.A., Bassey, U., Alba, K., Hamatani, T., Kargul, G.J., Luo, A.G., Kelso, J., Hide, W. and Ko, M.S.H.
 TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse cDNA clone set
 JOURNAL Genome Res. 12 (12), 1999-2003 (2002)
 MEDLINE 22354164

PUBMED
12466305
Other ESTs: H4074C02-3
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Casseil Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@gsun.grc.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://lgsun.grc.nia.nih.gov/cdna/NIA_7_4k.html for details.
Plate: H4074 row: C column: 02
Seq primer: -21M13 Reverse
High quality sequence stop: 479
POLYA=No.

FEATURES
source
Location/Qualifiers
1..479
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="H4074C02"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/clone_lib="NIA Mouse 7.4k cDNA Clone Set"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."

BASE COUNT
110 a 117 c 95 g 157 t

Query Match 1.2%; Score 22; DB 13; Length 479;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1560 AAAAAGTGTGATGAAATGCT 1581
DB 61 AAAAAGTGTGATGAAATGCT 40

RESULT 9
BH787917/c 495 bp DNA linear GSS 28-MAR-2002
LOCUS fmb020f010g10k0 fmb filtered library Zea mays genomic clone
DEFINITION fmb020f010g10 5', genomic survey sequence.
BH787917
VERSION BH787917.1 GI:19795777
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 495)
Bridman,M.A., Freese,R.G., Bedell,J.A., Nunberg,A.N. and Lakey,N.D.
Generetresh methylation filtered genomic sequences from maize
Unpublished
Contact: Bedell JA
Orion Genomics, LLC
4041 Forest Park Ave, St. Louis, MO 63108, USA
Tel: 314 615 6979
Fax: 314 615 5975
Email: jbedell@oriongenomics.com
Plate: fmb020f010 row: 5 column: 10
Seq primer: SK reverse
Class: shotgun
High quality sequence stop: 495.

FEATURES
source
Location/Qualifiers
1..495
/organism="Zea mays"
/mol_type="genomic DNA"
/cultivar="MO17"
/db_xref="taxon:4577"

/clone="fmb020f010g10"
/clone_lib="fmb filtered library"
/note="Organ: leaf; Vector: pBOSK(-); Site_1: HincII; DNA
prepared from purified nuclei was randomly sheared
end-repaired, size fractionated to enrich for the 0.5 to
5 kb fraction, ligated into HincII-digested pBOSK(-)
vector and electroporated into E. coli cells."
BASE COUNT 154 a 102 c 79 g 160 t

Query Match 1.2%; Score 22; DB 26; Length 495;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 750 GAAGGAATGTGACGTACCTT 771
DB 404 GAAGGAATGTGACGTACCTT 383

RESULT 10
BF198774 592 bp mRNA linear EST 03-NOV-2000
LOCUS 248732 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
DEFINITION BF198774
ACCESSION BF198774.1 GI:11090224
VERSION EST.
KEYWORDS Sus scrofa (pig)
SOURCE Sus scrofa
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
1 (bases 1 to 592)
Fahrenkrug,S.C., Smith,T.P.L., Freking,B.A., Cho,Y., White,J.,
Vallet,J., Wise,T., Rohrer,G.A., Petrea,G., Sultana,R., Quackenbush
J., and Keeler,J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
1226715
12233789

JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt-trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTCCACAGTCACGACG
Plate: 80 row: 1 column: 5
Seq primer: ATTGAGTGACACTATAG.
Location/Qualifiers
1..592
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_lib="MARC 2P1G"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT 179 a 88 c 130 g 195 t

Query Match 1.2%; Score 22; DB 10; Length 592;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1384 TTCTCTAGCTTTAGAGTTAA 1405
TTTTCTCTAGCTTTAGAGTTAA 1405

Db 39 TTTCTATGTTTGAAGTTAA 60

RESULT 11
B2421643
LOCUS

DEFINITION B2421643 653 bp DNA linear GSS 10-DEC-2002
h232f07.b1 WGS-bicolor (DH5a methyl filtered) Sorghum bicolor
Genomic clone h232f07 5', genomic survey sequence.

ACCESSION B2421643
VERSION B2421643.1 GI:26369150

KEYWORDS
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 653)
Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedhia, N.,
Katzemburger, F., King, L., Miller, B., Nascimben, L.,
Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)
Unpublished

AUTHORS

TITLE JOURNAL
COMMENT Contact: W. Richard McCombie
Lita Annenberg Hazen Genome Sequencing Center
Cold Spring Harbor Laboratory
PO Box 100, Cold Spring Harbor, NY 11724, USA
Tel: 516 367 8884
Fax: 516 367 8874
Email: mcombie@cshl.org
Plate: h232 row: f column: 07
Seq primer: -21M13UnivFwd
Class: shotgun
High quality sequence stop: 653.

FEATURES
source
1..653
location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="genomic DNA"
/db_xref="taxon:4558"
/clone="h232f07"
/lab_host="DH5a"
/note="Site 1: Xba I; Site 2: Xba I; The vector was
digested with XbaI and one nucleotide was added by fill in
in the recessive 3' end. The genomic DNA was methylated
end repaired, adaptor ligated and size fractionated using
sephadex. The resulting fragments were between 0.8 and 3
kb and were cloned into the vector (x/y reads in M13mp19,
.b/g reads in pUC19). The same ligation was transformed
into DH5a."

BASE COUNT 190 a 135 c 121 g 206 t 1 others
ORIGIN

Query Match 1.2%; Score 22; DB 29; Length 653;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 AGTCTGAACCTGTCATATTC 378
|||||
Db 289 AGTCTGAACCTGTCATATTC 310

RESULT 12
CA995790 661 bp mRNA linear EST 07-JAN-2003
LOCUS CA995790
DEFINITION r966a11.y1 Meloidogyne hapla c2 pAMP1 v1 Meloidogyne hapla cDNA 5',
mRNA sequence.
ACCESSION CA995790
VERSION CA995790.1 GI:27540661
KEYWORDS EST.
SOURCE Meloidogyne hapla
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.

REFERENCE
AUTHORS

1 (bases 1 to 661)
McCarter, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Maira, M., Hillier, L., Kuabla, T., Theising, B., Bowers, Y.,
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, R.,
Romko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and
Wilson, R.
The Washington Univ. Nematode EST Project, 1999
Unpublished

TITLE JOURNAL
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

The library was constructed by Claire Murphy and Dr. James McCarter
at Washington University, St. Louis. J2 were provided by Dr.
Valerie Williamson of the University of California at Davis
(vwilliamson@ucdavis.edu).
Seq primer: -40RP from Gibco
High quality sequence stop: 418.

FEATURES
source

1..661
location/Qualifiers
/organism="Meloidogyne hapla"
/mol_type="mRNA"
/db_xref="taxon:6305"
/dev_stage="J2"
/lab_host="DH10B"
/note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI;
The library was constructed by Claire Murphy and Dr. James
McCarter at Washington University, St. Louis. The cDNA was
made by using Dynabead Oligo-dT priming (Dyna). PCR based
library using a modified protocol from the SMART PCR cDNA
Synthesis Kit from Clontech. Directionally cloned into the
UDG sites of pAMP1. J2 were provided by Dr. Valerie
Williamson of the University of California at Davis
(vwilliamson@ucdavis.edu)."

BASE COUNT 232 a 75 c 67 g 287 t

Query Match 1.2%; Score 22; DB 14; Length 661;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1243 TTATTTAAATTTTACATTAAT 1264
|||||
Db 504 TTATTTAAATTTTACATTAAT 525

RESULT 13
CC283036/c 1136 bp DNA linear GSS 13-MAY-2003
LOCUS CC283036
DEFINITION CH261-112N11.Sp6.1 CH261 Gallus gallus genomic clone CH261-112N11,
genomic survey sequence.
ACCESSION CC283036
VERSION CC283036.1 GI:30649133
KEYWORDS GSS.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 1136)
Kremetzki, C., Higsbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Wards, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished
Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine

Email: submissions@watson.wustl.edu
Insert Length: 18200 Std Error: 0.00
Seq primer: SP6 ATTAGGTGACACTATG
Class: BAC ends
High quality sequence start: 73
High quality sequence stop: 682.

FEATURES

source

1. 1136

/organism="Gallus gallus"

/mol_type="Genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="taxon:9031"

/clone="CH261-112N11"

/sex="female"

/cell_line="UCD001, indred 256"

/clone_lib="CH261"

/note="Vector: pPARAC2.1; Site 1: EcoRI; Site 2: EcoRI;

CH261 Female Chicken library - for library and clone

ordering information: <http://www.chori.org/bacpac>"

BASE COUNT 338 a 208 c 174 g 415 t 1 others

ORIGIN

Query Match 1.2% Score 22; DB 29; Length 1136;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

388 AGAAAAATTTCTAGTGTGTTAG 409
894 AGAAAAATTTCTAGTGTGTTAG 873

RESULT 14

AU060103/c 196 bp mRNA linear EST 20-MAY-1999
LOCUS AU060103 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
DEFINITION AU060103 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
ACCESSION AU060103
VERSION AU060103.1 GI:4881207

KEYWORDS

EST; Dictyostelium discoideum
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.

REFERENCE

AUTHORS

Morito, T., Urushihara, H., Satoh, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B. N., Pi, M., Satoh, T., Takekoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.

Developmental cDNA in Dictyostelium discoideum
Unpublished

JOURNAL

Institute of Biological Sciences
University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan
Tel: 81-298-53-4664
Fax: 81-298-53-6614

Email: hideo@biol.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers
1. 1396

/organism="Dictyostelium discoideum"

/mol_type="mRNA"

/strain="XX4"

/db_xref="taxon:44689"

/clone="SLA320"

/dev_stage="slug"

/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"

BASE COUNT 98 a 36 c 15 g 47 t

ORIGIN

Query Match 1.2% Score 21; DB 9; Length 196;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1032 ATTTTGAATTATGTTTCT 1052

Db 66 ATTTTGAATTATGTTTCT 46

RESULT 15
BO851931/c 272 bp mRNA linear EST 14-AUG-2002
LOCUS BO851931
DEFINITION BO851931, YG.ab1 OG ABCDI lettuce salinas Lactuca sativa cDNA clone
OG816K21, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Lactuca sativa

Search completed: November 25, 2003, 19:00:34
Job time : 2621 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 18:16:43 ; Search time 440 Seconds
(without alignments)

13350.983 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atcaacttcgactcttga.....tagagcgatcaagctgaacc 1783

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 219069 seqs, 1647345023 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PTCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PTCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|----------------------|
| 1 | 22 | 1.2 | 1067 | 12 | US-10-027-632-250060 |
| 2 | 22 | 1.2 | 1067 | 13 | US-10-027-632-250060 |
| 3 | 21 | 1.2 | 7400 | 12 | US-09-960-858-4 |
| 4 | 21 | 1.2 | 7400 | 12 | US-09-960-858-4 |
| 5 | 21 | 1.2 | 8011 | 12 | US-10-311-455-52 |
| 6 | 21 | 1.2 | 580073 | 12 | US-10-025-220-1 |
| 7 | 20 | 1.1 | 605 | 12 | US-10-027-632-210764 |
| 8 | 20 | 1.1 | 605 | 13 | US-10-027-632-210764 |
| 9 | 20 | 1.1 | 663 | 12 | US-10-027-632-180487 |
| 10 | 20 | 1.1 | 663 | 12 | US-10-027-632-180487 |
| 11 | 20 | 1.1 | 663 | 13 | US-10-027-632-180487 |
| 12 | 20 | 1.1 | 663 | 13 | US-10-027-632-180487 |
| 13 | 20 | 1.1 | 1738 | 14 | US-10-204-887-67 |
| 14 | 20 | 1.1 | 2000 | 10 | US-09-938-842A-4813 |
| 15 | 20 | 1.1 | 5686 | 9 | US-09-775-938A-31 |
| 16 | 20 | 1.1 | 7351 | 12 | US-10-311-455-1 |

| | | | | | | |
|----|----|-----|---------|----|----------------------|--------------------|
| 17 | 20 | 1.1 | 16950 | 14 | US-10-114-170-166 | Sequence 166, App |
| 18 | 20 | 1.1 | 42611 | 12 | US-10-161-127-2 | Sequence 2, Appl |
| 19 | 20 | 1.1 | 326014 | 9 | US-09-731-231A-3 | Sequence 3, Appl |
| 20 | 20 | 1.1 | 3673778 | 12 | US-10-312-841-1 | Sequence 1, Appl |
| 21 | 19 | 1.1 | 394 | 8 | US-08-781-986A-1543 | Sequence 1943, Ap |
| 22 | 19 | 1.1 | 585 | 12 | US-10-027-632-90502 | Sequence 90502, A |
| 23 | 19 | 1.1 | 585 | 13 | US-10-027-632-90502 | Sequence 90502, A |
| 24 | 19 | 1.1 | 598 | 12 | US-10-027-632-210749 | Sequence 210749, |
| 25 | 19 | 1.1 | 598 | 13 | US-10-027-632-210749 | Sequence 210749, |
| 26 | 19 | 1.1 | 644 | 12 | US-10-027-632-190385 | Sequence 190385, |
| 27 | 19 | 1.1 | 644 | 13 | US-10-027-632-190385 | Sequence 190385, |
| 28 | 19 | 1.1 | 728 | 12 | US-10-027-632-149996 | Sequence 149996, |
| 29 | 19 | 1.1 | 728 | 13 | US-10-027-632-149996 | Sequence 149996, |
| 30 | 19 | 1.1 | 780 | 12 | US-10-027-632-149997 | Sequence 149997, |
| 31 | 19 | 1.1 | 780 | 13 | US-10-027-632-149997 | Sequence 149997, |
| 32 | 19 | 1.1 | 813 | 9 | US-09-770-445-763 | Sequence 763, App |
| 33 | 19 | 1.1 | 826 | 14 | US-10-091-572-898 | Sequence 898, App |
| 34 | 19 | 1.1 | 826 | 14 | US-10-091-572-898 | Sequence 900, App |
| 35 | 19 | 1.1 | 1500 | 10 | US-09-822-830A-423 | Sequence 423, App |
| 36 | 19 | 1.1 | 1507 | 12 | US-10-027-632-264404 | Sequence 264404, |
| 37 | 19 | 1.1 | 1507 | 13 | US-10-027-632-264404 | Sequence 264404, |
| 38 | 19 | 1.1 | 5269 | 12 | US-10-311-455-2030 | Sequence 2030, Ap |
| 39 | 19 | 1.1 | 5347 | 12 | US-10-311-455-1641 | Sequence 1641, App |
| 40 | 19 | 1.1 | 5930 | 12 | US-10-311-455-490 | Sequence 490, App |
| 41 | 19 | 1.1 | 6070 | 12 | US-10-204-708-9 | Sequence 9, Appl |
| 42 | 19 | 1.1 | 6070 | 12 | US-10-311-455-213 | Sequence 213, App |
| 43 | 19 | 1.1 | 6125 | 12 | US-10-311-455-1585 | Sequence 1585, Ap |
| 44 | 19 | 1.1 | 6125 | 12 | US-10-240-453-151 | Sequence 151, App |
| 45 | 19 | 1.1 | 6341 | 12 | US-10-311-455-1617 | Sequence 1617, Ap |

ALIGNMENTS

RESULT 1
US-10-027-632-250060
; Sequence 250060, Application US/10027632
; Publication No. US20030204075A5
; GENERAL INFORMATION:
; APPLICANT Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027, 632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 250060
; LENGTH: 1067
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250060
Query Match 1.2%; Score 22; DB 12; Length 1067;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1037 TGAATTATGTTTCATGTA 1058
|||||

Db 305 TGAATTATGTTTCTATGTTA 326

RESULT 2

US-10-027-632-250060
; Sequence 250060, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250060
; LENGTH: 1067
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250060

Query Match 1.2%; Score 22; DB 13; Length 1067;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 TGAATTATGTTTCTATGTTA 1058

Db 305 TGAATTATGTTTCTATGTTA 326

RESULT 3

US-09-960-858-4/c
; Sequence 4, Application US/09960858
; Publication No. US2003013877A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
; FILE OF INVENTION: USE
; FILE REFERENCE: P-EA 4974
; CURRENT APPLICATION NUMBER: US/09/960,858
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7400
; TYPE: DNA
; ORGANISM: M. genitalium
US-09-960-858-4

Query Match 1.2%; Score 21; DB 12; Length 7400;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGTTTAACTTGT 1492

Db 6848 AACATTGTTTAACTTGT 6828

RESULT 4

US-09-960-870-4/c

; Sequence 4, Application US/09960870
; Publication No. US20030134281A1
; GENERAL INFORMATION:
; APPLICANT: Evans, Glen
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF
; FILE OF INVENTION: USE
; FILE REFERENCE: P-EA 4738
; CURRENT APPLICATION NUMBER: US/09/960,870
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 7400
; TYPE: DNA
; ORGANISM: M. genitalium
US-09-960-870-4

Query Match 1.2%; Score 21; DB 12; Length 7400;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGTTTAACTTGT 1492

Db 6848 AACATTGTTTAACTTGT 6828

RESULT 5

US-10-311-455-52
; Sequence 52, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 52
; LENGTH: 8011
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-52

Query Match 1.2%; Score 21; DB 12; Length 8011;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTTATATTGTGT 1379

Db 5456 ATATTATTTATATTGTGT 5476

RESULT 6

US-10-205-220-1
; Sequence 1, Application US/10205220
; Publication No. US20030170663A1
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragmen
; FILE REFERENCE: PB193PDI

```

; CURRENT APPLICATION NUMBER: US/10/205,220
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 08/545,528
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-10-205-220-1
```

```

Query Match          1.2%; Score 21; DB 12; Length 580073;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1472 AACATTGGTTTAACTTGT 1492
      |||||||
Db      494176 AACATTGGTTTAACTTGT 494196
```

```

RESULT 7
US-10-027-632-210764
; Sequence 210764, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210764
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-210764
```

```

Query Match          1.1%; Score 20; DB 12; Length 605;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1036 TTGAATTATGTTTCTATG 1055
      |||||||
Db      240 TTGAATTATGTTTCTATG 259
```

```

RESULT 8
US-10-027-632-210764
; Sequence 210764, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210764
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-210764
```

```

Query Match          1.1%; Score 20; DB 13; Length 605;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1036 TTGAATTATGTTTCTATG 1055
      |||||||
Db      240 TTGAATTATGTTTCTATG 259
```

```

RESULT 9
US-10-027-632-180487
; Sequence 180487, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180487
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-180487
```

```

Query Match          1.1%; Score 20; DB 12; Length 663;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1204 CTCCTATTTTAAATTTTACA 1223
```

Db 550 CTTTATTTTAAATTTCACA 569

```

RESULT 10
US-10-027-632-180488
; Sequence 180488, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180488
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-180488

Query Match 1.1%; Score 20; DB 12; Length 663;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1204 CTTTATTTTAAATTTCACA 1223

```

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 11
US-10-027-632-180487
; Sequence 180487, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180487

```

LENGTH: 663
TYPE: DNA
ORGANISM: Human
US-10-027-632-180487

Query Match 1.1%; Score 20; DB 13; Length 663;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 CTTTATTTTAAATTTCACA 1223

```

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 12
US-10-027-632-180488
; Sequence 180488, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/199,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180488
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-180488

```

Query Match 1.1%; Score 20; DB 13; Length 663;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 CTTTATTTTAAATTTCACA 1223

```

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 13
US-10-204-887-67/c
; Sequence 67, Application US/10204887
; Publication No. US20030124569A1
; GENERAL INFORMATION:
; APPLICANT: INCTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.

```



```

; APPLICANT: DUFOR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: RUSSO, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFEO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1134 FCT
; CURRENT APPLICATION NUMBER: US/10/204,887
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
; 2000-05-17; 2000-05-17
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 67
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No. US20030124569A1 LI:336872.1:2000MAY01
US-10-204-887-67

Query Match 1.1%; Score 20; DB 14; Length 1738;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TCTCTCAATATATCTGATAT 1362
DB 620 TCTCTCAATATATCTGATAT 601

RESULT 14
US-09-938-842A-4813/c
; Sequence 4813, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4813
; LENGTH: 2000
; TYPE: DNA

```

```

; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4813

Query Match 1.1%; Score 20; DB 10; Length 2000;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 AAAGATTATTAAGTTGAAT 238
DB 558 AAAGATTATTAAGTTGAAT 539

RESULT 15
US-09-775-938A-31
; Sequence 31, Application US/09775938A
; Patent No. US2002008165A1
; GENERAL INFORMATION:
; APPLICANT: Haygood, M.
; APPLICANT: Davidson, S.K.
; APPLICANT: Allen, S.W.
; APPLICANT: Hildebrand, M.
; TITLE OF INVENTION: Bryostatins, Bryopyrans and Polyketides: Compositions and Method
; FILE REFERENCE: 1133.010U1
; CURRENT APPLICATION NUMBER: US/09/775,938A
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US00/21326
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/147,283
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 5686
; TYPE: DNA
; ORGANISM: Endobugula sertula
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(5686)
; OTHER INFORMATION: N refers to any nucleotide.
US-09-775-938A-31

Query Match 1.1%; Score 20; DB 9; Length 5686;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1052 TATGTTACTTTCTTCAGC 1071
DB 4061 TATGTTACTTTCTTCAGC 4060

```

Search completed: November 25, 2003, 20:19:30
 Job time : 444 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 16:36:37 ; Search time 95 Seconds

(without alignments)
8284.064 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccactctctgactcttga.....tagagcgatcaagctgacac 1783

Scoring table: OLIGO NUC

Gapop 60.0 / Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCUS COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------------------|-------------------|
| 1 | 21 | 1.2 | 580073 | 4 US-08-545-528D-1 | Sequence 1, Appl |
| 2 | 20 | 1.1 | 1500 | 4 US-09-107-532A-1870 | Sequence 1870, Ap |
| 3 | 20 | 1.1 | 16950 | 4 US-09-453-702B-166 | Sequence 166, App |
| 4 | 19 | 1.1 | 1882 | 1 US-08-257-073-12 | Sequence 12, Appl |
| 5 | 19 | 1.1 | 1884 | 1 US-08-257-073-8 | Sequence 8, Appl |
| 6 | 18 | 1.0 | 233 | 2 US-08-687-080-70 | Sequence 70, Appl |
| 7 | 18 | 1.0 | 741 | 4 US-09-328-352-3895 | Sequence 3895, Ap |
| 8 | 18 | 1.0 | 1042 | 3 US-08-765-381-3 | Sequence 3, Appl |
| 9 | 18 | 1.0 | 1079 | 3 US-08-765-381-1 | Sequence 1, Appl |
| 10 | 18 | 1.0 | 1303 | 2 US-08-793-410-11 | Sequence 11, Appl |
| 11 | 18 | 1.0 | 1947 | 4 US-09-134-001C-2324 | Sequence 2324, Ap |
| 12 | 18 | 1.0 | 2186 | 4 US-08-961-527-179 | Sequence 179, App |
| 13 | 18 | 1.0 | 3001 | 4 US-09-539-333D-184 | Sequence 184, App |
| 14 | 18 | 1.0 | 83450 | 4 US-09-811-469-3 | Sequence 3, Appl |
| 15 | 18 | 1.0 | 580073 | 4 US-08-545-528D-1 | Sequence 1, Appl |
| 16 | 17 | 1.0 | 40 | 1 US-08-199-507B-26 | Sequence 26, Appl |
| 17 | 17 | 1.0 | 40 | 1 US-08-441-828-26 | Sequence 26, Appl |
| 18 | 17 | 1.0 | 318 | 3 US-09-018-584A-10 | Sequence 10, Appl |
| 19 | 17 | 1.0 | 363 | 4 US-09-601-198-169 | Sequence 169, App |
| 20 | 17 | 1.0 | 471 | 4 US-09-134-001C-1423 | Sequence 1423, Ap |
| 21 | 17 | 1.0 | 689 | 4 US-09-149-476-263 | Sequence 263, App |
| 22 | 17 | 1.0 | 743 | 4 US-09-220-132-37 | Sequence 37, Appl |
| 23 | 17 | 1.0 | 1001 | 4 US-09-641-638-133 | Sequence 133, App |
| 24 | 17 | 1.0 | 1001 | 4 US-09-641-638-134 | Sequence 134, App |
| 25 | 17 | 1.0 | 1282 | 3 US-09-475-316A-69 | Sequence 69, Appl |
| 26 | 17 | 1.0 | 1330 | 4 US-09-134-001C-2066 | Sequence 2066, Ap |
| 27 | 17 | 1.0 | 1353 | 4 US-09-252-991A-7869 | Sequence 7869, Ap |

| | | | | | |
|----|----|-----|------|-----------------------|-------------------|
| 28 | 17 | 1.0 | 1398 | 4 US-09-134-001C-1019 | Sequence 1019, Ap |
| 29 | 17 | 1.0 | 1620 | 4 US-09-328-352-1307 | Sequence 1307, Ap |
| 30 | 17 | 1.0 | 1751 | 4 US-09-149-476-110 | Sequence 110, App |
| 31 | 17 | 1.0 | 1939 | 1 US-07-715-751B-2 | Sequence 2, Appl |
| 32 | 17 | 1.0 | 2178 | 3 US-09-334-601-6 | Sequence 6, Appl |
| 33 | 17 | 1.0 | 2288 | 3 US-09-334-601-1 | Sequence 1, Appl |
| 34 | 17 | 1.0 | 2359 | 4 US-09-425-488-7 | Sequence 7, Appl |
| 35 | 17 | 1.0 | 2388 | 3 US-09-276-531-89 | Sequence 89, Appl |
| 36 | 17 | 1.0 | 2404 | 1 US-08-484-101B-43 | Sequence 43, Appl |
| 37 | 17 | 1.0 | 2404 | 3 US-08-714-524D-43 | Sequence 43, Appl |
| 38 | 17 | 1.0 | 2445 | 3 US-09-298-367B-9 | Sequence 9, Appl |
| 39 | 17 | 1.0 | 2703 | 4 US-09-620-312D-152 | Sequence 152, App |
| 40 | 17 | 1.0 | 2883 | 2 US-08-533-306A-7 | Sequence 7, Appl |
| 41 | 17 | 1.0 | 2883 | 2 US-08-742-923A-7 | Sequence 7, Appl |
| 42 | 17 | 1.0 | 2992 | 2 US-08-841-349-10 | Sequence 10, Appl |
| 43 | 17 | 1.0 | 3267 | 4 US-09-328-352-1597 | Sequence 1597, Ap |
| 44 | 17 | 1.0 | 3494 | 3 US-09-334-601-5 | Sequence 5, Appl |
| 45 | 17 | 1.0 | 3729 | 4 US-09-107-532A-1587 | Sequence 1587, Ap |

ALIGNMENTS

RESULT 1
US-08-545-528D-1
Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:
APPLICANT: Fraser et al.
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
Patent No. 6537773
FILE REFERENCE: PB193P1
CURRENT APPLICATION NUMBER: US/08/545,528D
CURRENT FILING DATE: 1995-10-19
PRIOR APPLICATION NUMBER: US 08/468,018
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/473,545
PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 580073
TYPE: DNA
ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 1.2% Score 21; DB 4; Length 580073;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGGTTTAACTTGTT 1492
DB 494176 AACATTGGTTTAACTTGTT 494196

RESULT 2
US-09-107-532A-1870
Sequence 1870, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stramm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FABRICUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: JULY 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Denise
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-8277
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 1870:
SEQUENCE CHARACTERISTICS:
LENGTH: 1500 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1..1500
SEQUENCE DESCRIPTION: SEQ ID NO: 1870:
US-09-107-532A-1870
Query Match 1.1%; Score 20; DB 4; Length 1500;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 619 GGAAGATCTTATGAGAAAT 638
DB 288 GGAAGATCTTATGAGAAAT 307
RESULT 3
US-09-453-702B-166
Sequence 166, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Perna, Nicole T.
Blunkett, Guy
Welch, Rod
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Plunkney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960236.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 16950
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-453-702B-166
Query Match 1.1%; Score 20; DB 4; Length 16950;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1204 CTTTATTAAATTGACA 1223
DB 2917 CTTTATTAAATTGACA 2936
RESULT 4
US-08-257-073-12/C
Sequence 12, Application US/08257073
Patent No. 5765597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
APPLICANT: Tine, John A.
TITLE OF INVENTION: MALARIA RECOMBINANT POXYVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 1882 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-12

Query Match 1.1%; Score 19; DB 1; Length 1882;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1705 TTTTGGCATATATTCGCTC 1723
DB 348 TTTTGGCATATATTCGCTC 330

RESULT 5
US-08-257-073-8/c
Sequence 8, Application US/08257073
Patent No. 5766597
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
APPLICANT: de Taisne, Charles
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
NUMBER OF SEQUENCES: 143
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTWS
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1884 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-8

Query Match 1.1%; Score 19; DB 1; Length 1884;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1705 TTTTGGCATATATTCGCTC 1723
DB 348 TTTTGGCATATATTCGCTC 330

DB 348 TTTTGGCATATATTCGCTC 330

RESULT 6
US-08-687-080-70
Sequence 70, Application US/08687080
Patent No. 5965427
GENERAL INFORMATION:

APPLICANT: Gregory Dolganov
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
NUMBER OF SEQUENCES: 175
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholez, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 233 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: INTRON 6 OF RAD50 GENOMIC SEQUENCE
US-08-687-080-70

Query Match 1.0%; Score 18; DB 2; Length 233;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 TCATATTTCTTATGTTT 1396
DB 182 TCATATTTCTTATGTTT 199

RESULT 7
US-09-328-352-3895
Sequence 3895, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Brelton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAWMANNI FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-032A
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3895
LENGTH: 741

TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-3895

Query Match 1.0%; Score 18; DB 4; Length 741;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1217 TTTTACACATATATATAA 1234
DB 62 TTTTACACATATATATAA 79

RESULT 8
US-08-765-381-3
Sequence 3, Application US/08765381
Patent No. 6083724
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
TITLE OF INVENTION: sequences encoding same
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSES: Scully Scott Murphy and Presser
STREET: 400 Garden City Plaza
CITY: Garden City, New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,381
FILING DATE: 19-DEC-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN1542/95
FILING DATE: 06-MAR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00114
FILING DATE: 05-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESSER, LEOPOLD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-516-742-4343
TELEFAX: 1-516-742-4366
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1042 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Chicken (gallus sp.)
FEATURE:
NAME/KEY: CDS
LOCATION: 97..588
US-08-765-381-3

Query Match 1.0%; Score 18; DB 3; Length 1042;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1244 TATTTAAATTTTACAAT 1261
DB 620 TATTTAAATTTTACAAT 637

RESULT 9
US-08-765-381-1
Sequence 1, Application US/08765381
Patent No. 6083724
GENERAL INFORMATION:
APPLICANT: Commonwealth Scientific and Industrial Research Organisation
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic
TITLE OF INVENTION: sequences encoding same
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESSES:
ADDRESSES: Scully Scott Murphy and Presser
STREET: 400 Garden City Plaza
CITY: Garden City, New York
STATE: New York
COUNTRY: UNITED STATES OF AMERICA
ZIP: 11530

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/765,381
FILING DATE: 19-DEC-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: AU PN1542/95
FILING DATE: 06-MAR-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/AU96/00114
FILING DATE: 05-MAR-1996
ATTORNEY/AGENT INFORMATION:
NAME: PRESSER, LEOPOLD
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-516-742-4343
TELEFAX: 1-516-742-4366
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Chicken (gallus sp.)
CELL TYPE: T-cell
CELL LINE: CC8.1h
IMMEDIATE SOURCE:
LIBRARY: CC8.1h
CLONE: ChIFN-gamma
FEATURE:
NAME/KEY: CDS
LOCATION: 134..625
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 191..625
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..133
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 626..1079
US-08-765-381-1

Query Match 1.0%; Score 18; DB 3; Length 1079;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1244 TATTTAAATTTTACAAT 1261
DB 657 TATTTAAATTTTACAAT 674

RESULT 10

```

US-08-793-410-11/c
/ Sequence 11, Application US/08793410
/ Patent No. 5955650
/ GENERAL INFORMATION:
/ APPLICANT: HITZ, WILLIAM DEAN
/ TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA
/ TITLE OF INVENTION: AND SOYBEAN PALMITOYL-ACP THIO-
/ TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN
/ TITLE OF INVENTION: THE REGULATION OF FATTY ACID
/ TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN
/ NUMBER OF INVENTION: 32
/ NUMBER OF SEQUENCES: 32
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
/ STREET: 1007 MARKET STREET
/ CITY: WILMINGTON
/ STATE: DELAWARE
/ COUNTRY: USA
/ ZIP: 19898
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: DISKETTE, 3.50 INCH
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: MICROSOFT WINDOWS 95
/ SOFTWARE: MICROSOFT WORD VERSION 7.0A
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/793,410
/ FILING DATE:
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/10627
/ FILING DATE: AUGUST 25, 1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: CHRISTENBURY, LYNN M.
/ REGISTRATION NUMBER: 30,971
/ REFERENCE/DOCKET NUMBER: CR-9567-A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 302-992-5481
/ TELEFAX: 302-773-0164
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1303 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ US-08-793-410-11

Query Match 1.0%; Score 18; DB 2; Length 1303;
Best local Similarity 100.0%; Pred. NO. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 403 TGTTAGAATTTGACTT 420
|||
Db 577 TGTTAGAATTTGACTT 560

RESULT 11
US-09-134-001C-2324/c
/ Sequence 2324, Application US/09134001C
/ Patent No. 6380370
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
/ TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ PRIOR FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/055,779
/ PRIOR FILING DATE: 1997-08-14

```

```

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2324
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2324

Query Match          1.0%; Score 18; Length 1947;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      544 GTGCATTGACCATTTTG 561
      |||||||
DB      1747 GTGCATTGACCATTTTG 1730

RESULT 12
US-08-961-527-179/c
; Sequence 179; Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 179:
SEQUENCE CHARACTERISTICS:
; LENGTH: 2186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-179

Query Match          1.0%; Score 18; DB 4; Length 2186;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      394 ATTCTAAGCTTAGAA 411
      |||||||
DB      466 ATTCTAAGCTTAGAA 449

RESULT 13
US-09-539-333D-184/c
; Sequence 184; Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel

```

```

; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguenel, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 184
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; LOCATION: 1501
; OTHER INFORMATION: 99-25906-131 : polymorphic base G or T
; NAME/KEY: misc_binding
; LOCATION: 1481..1500
; OTHER INFORMATION: 99-25906-131.misl,
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1502..1520
; OTHER INFORMATION: 99-25906-131.misl2, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1374..1392
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1888..1908
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-25906-131 probe
; US-09-539-333D-184

Query Match          1.0%; Score 18; DB 4; Length 3001;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1133 AGAATATGATGATTTC 1150
Db      778 AGAATATGATGATTTC 761

```

```

RESULT 14
US-09-811-469-3/c
; Sequence 3, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:

```

```

; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: C0001171
; CURRENT APPLICATION NUMBER: US/09/811,469
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 83450
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(83450)
; OTHER INFORMATION: n = A,T,C or G
; US-09-811-469-3

```

```

Query Match          1.0%; Score 18; DB 4; Length 83450;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1422 AAAGTAAATTCATAT 1439
Db      15264 AAAGTAAATTCATAT 15247

```

```

RESULT 15
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide sequence of the Mycoplasma Genitalium Genome, Fragment
; FILE REFERENCE: P019391
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
; US-08-545-528D-1

```

```

Query Match          1.0%; Score 18; DB 4; Length 580073;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1744 TAAACATGACTAATT 1761
Db      574359 TAAACATGACTAATT 574342

```

```

Search completed: November 25, 2003, 19:02:29
Job time : 98 secs

```


PI Heim U, Weber H;
XX WPI: 2000-365631/31.
DR P-PSDB; AAB10028.
XX
PT Expression cassette for expressing genes in plant seeds, useful for
PT producing enzymes or pharmaceuticals, includes the promoter from a
PT sucrose-binding protein-related gene
PS
PS Disclosure; Fig 1; 24pp; German.

XX
XX This invention describes a novel expression cassette (I) for expressing
CC genes in plant seeds which comprises (i) the promoter of the
CC sucrose-binding protein (SBP)-like seed protein (II), (iii) optionally the
CC sequence for a signal peptide, particularly from (II), (iii) the gene
CC (iii) to be expressed and (iv) a 3'-termination sequence. (I) and a
CC plasmid containing the expression cassette (IV) are used for the
CC expression of homologous or heterologous genes in the seeds of
CC transformed plants, particularly genes that alter the storage properties
CC and germination capacity of the seeds. Alternatively, transgenic plants
CC that express altered or new products in their seeds are selected, grown
CC to establish stable lines and the resulting products (e.g. enzymes,
CC pharmaceuticals or proteins that contain essential amino acids) are
CC extracted. Also insertion of (iii) in the antisense orientation may be
CC used to reduce or switch off expression of particular genes. (I) provide
CC seed-specific expression (in cotyledons and endosperm) with stable
CC expression at high level. They provide an overall increase in expression
CC rate, improve utilization of the developmental period of the seed and
CC can overcome the effect of co-suppression. This sequence encodes a
CC fragment of the Vicia faba (soybean) sucrose binding protein (SBP) which
CC is described in the method of the invention.
CC
XX

SO Sequence 1909 BP; 638 A; 317 C; 281 G; 673 T; 0 other;

Query Match 100.0%; Score 1783; DB 21; Length 1909;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCCAGCTTCGATCTTGAATCTCTGTCGCAACATGTTGAGAGATTCTAAGACT 60
DB 1 ATCCAGCTTCGATCTTGAATCTCTGTCGCAACATGTTGAGAGATTCTAAGACT 60
QY 61 TTTCAGAAAGCTTTGAAATGCTTTGAGACTTTCTTGAATTACTCTTGCAGAACTCTGA 120
DB 61 TTTCAGAAAGCTTTGAAATGCTTTGAGACTTTCTTGAATTACTCTTGCAGAACTCTGA 120
QY 121 TTGACCTAGCTGAAAGCTGCTCCAGAGTTCTAACCAATTCGCTTGGAGAGGCCA 180
DB 121 TTGACCTAGCTGAAAGCTGCTCCAGAGTTCTAACCAATTCGCTTGGAGAGGCCA 180
QY 121 TTGAACCTAGCTGAAAGCTGCTCCAGAGTTCTAACCAATTCGCTTGGAGAGGCCA 180
DB 121 TTGAACCTAGCTGAAAGCTGCTCCAGAGTTCTAACCAATTCGCTTGGAGAGGCCA 180
QY 181 AATTTATGAGTACTTCAGTTTCATGAGAGCTGTCTTCAAGATTTTAACTTGAATTC 240
DB 181 AATTTATGAGTACTTCAGTTTCATGAGAGCTGTCTTCAAGATTTTAACTTGAATTC 240
QY 241 CATCATTTTAAAGAGAGTTCTGTTCCGCAATGCTTGAATCTCATTTGAATCTCAACT 300
DB 241 CATCATTTTAAAGAGAGTTCTGTTCCGCAATGCTTGAATCTCATTTGAATCTCAACT 300
QY 301 CTGTGTGAGAGAGTTCTTCCAGATCAACTGATCATGTGGAATACTGGCCAGAGATT 360
DB 301 CTGTGTGAGAGAGTTCTTCCAGATCAACTGATCATGTGGAATACTGGCCAGAGATT 360
QY 361 CTGAACTGTGATTTCTTAACTGTAAGAAAATTTCTAAGTTTGAATTTGACTT 420
DB 361 CTGAACTGTGATTTCTTAACTGTAAGAAAATTTCTAAGTTTGAATTTGACTT 420
QY 421 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480
DB 421 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480
QY 481 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480
DB 481 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480
QY 481 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480
DB 481 TTCCGAAGCAACTGATCTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 480

QY 541 AGTGTGCAATGACCATTTTGTCTTGTGCAATTCGCAACCTTAATGATGATGATG 600
DB 541 AGTGTGCAATGACCATTTTGTCTTGTGCAATTCGCAACCTTAATGATGATGATG 600
QY 601 CTGCAAACTGATGTCATGAGAGATCTTATGAGAAAATTTCTGAGACTGAGAGAAA 660
DB 601 CTGCAAACTGATGTCATGAGAGATCTTATGAGAAAATTTCTGAGACTGAGAGAAA 660
QY 661 TTTTGTGATGACACAAAGAAATCTGTTTTCATGCGGACATGACATTAACATTA 720
DB 661 TTTTGTGATGACACAAAGAAATCTGTTTTCATGCGGACATGACATTAACATTA 720
QY 721 AACCCACTTATTCGAGAGAGTGAATGAGAGAAATGTCAGTTACTTTCGAGTT 780
DB 721 AACCCACTTATTCGAGAGAGTGAATGAGAGAAATGTCAGTTACTTTCGAGTT 780
QY 781 CATTAAGCAACTTACAGACCTTTTACTTAAATACTACAAAGAGAAATTTAAACAC 840
DB 781 CATTAAGCAACTTACAGACCTTTTACTTAAATACTACAAAGAGAAATTTAAACAC 840
QY 841 TTAGAGAGTATGGAGGTTAAAGAGACATTTAAGGGGAGTAAATTAATGTTG 900
DB 841 TTAGAGAGTATGGAGGTTAAAGAGACATTTAAGGGGAGTAAATTAATGTTG 900
QY 901 TTGTACCAACCACTACCTTTAGTATTAAGAAATTTGATATCATCATTAATAT 960
DB 901 TTGTACCAACCACTACCTTTAGTATTAAGAAATTTGATATCATCATTAATAT 960
QY 961 TATGTCTCTTATTTAAATTTGATTAAGTGTATGATGATGAGAAACCAATATG 1020
DB 961 TATGTCTCTTATTTAAATTTGATTAAGTGTATGATGATGAGAAACCAATATG 1020
QY 1021 TCTGTCTGATTTTGAATTTGTTTCTATGTTACTTTCTTCAAGCCATATATA 1080
DB 1021 TCTGTCTGATTTTGAATTTGTTTCTATGTTACTTTCTTCAAGCCATATATA 1080
QY 1081 AACTTTGATGCTAAATTTGATGCTGAGAAAAATGCTGATGATGATGATGATG 1140
DB 1081 AACTTTGATGCTAAATTTGATGCTGAGAAAAATGCTGATGATGATGATGATG 1140
QY 1141 TGGTATTTCAAGTCAAAATTCATCATTAAGAAATTTGATGATGATGATGATG 1200
DB 1141 TGGTATTTCAAGTCAAAATTCATCATTAAGAAATTTGATGATGATGATGATG 1200
QY 1201 ATTCTCTTATTTAAATTTTACACATTAATTAATTTCTTATTTTAAATTTTAA 1260
DB 1201 ATTCTCTTATTTAAATTTTACACATTAATTAATTTCTTATTTTAAATTTTAA 1260
QY 1261 TAAATTAATTTTACACCTGTCACCTTTGAGATGACACCAACATTTAATCTTA 1320
DB 1261 TAAATTAATTTTACACCTGTCACCTTTGAGATGACACCAACATTTAATCTTA 1320
QY 1321 TTTATTTCTTAATTTTGAATCTGCAATATATGATATTTTATTTATTTTGTGTC 1380
DB 1321 TTTATTTCTTAATTTTGAATCTGCAATATATGATATTTTATTTATTTTGTGTC 1380
QY 1381 AATTTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 1440
DB 1381 AATTTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAAT 1440
QY 1441 TGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 1500
DB 1441 TGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 1500
QY 1501 AAAAGTATTAATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1560
DB 1501 AAAAGTATTAATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1560
QY 1561 AAAAGTATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 1620
DB 1561 AAAAGTATTAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGA 1620

QY 1621 AATTCATTTCAGATGATAGAACTGCCATACGAATTAATTAGCATTAAGACAGTATG 1680
 DB 1621 AATTCATTTCAGATGATAGAACTGCCATACGAATTAATTAGCATTAAGACAGTATG 1680
 QY 1681 TTAACACACGTCCTCCCTTCGATGTTTTTGGCCATAATATTCGCTCTCTTTTCTTCAC 1740
 DB 1681 TTAACACACGTCCTCCCTTCGATGTTTTTGGCCATAATATTCGCTCTCTTTTCTTCAC 1740
 QY 1741 GATATTAACATGATGATTAATTAATAGAGCGGATCAAGCTGAACC 1783
 DB 1741 GATATTAACATGATGATTAATTAATAGAGCGGATCAAGCTGAACC 1783

RESULT 2
 ABL32079
 ID ABL32079 standard; DNA; 8011 BP.
 XX
 AC ABL32079;

26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 52.

Human; immune system disease; cytosine methylation; antiasthmatic;
 antiarteriosclerotic; antianaemic; cytosolic; noctropic;
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 antirheumatic; antiarthritic; antidiabetic; ophtalmological;
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 gene; ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002.

02-JUL-2001; 2001WO-EP07537.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-130909/17.

Nucleic acid comprising fragment of chemically modified gene, useful

for diagnosis and treatment of diseases associated with abnormal

cytosine methylation -

Claim 1; SEQ ID NO 52; 32bp + Sequence Listing; German.

The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

Sequence 8011 BP; 2439 A; 145 C; 1810 G; 3617 T; 0 other;

Query Match 1.2%; Score 21; DB 24; Length 8011;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTTTATTTATTTGCT 1379
 DB 5456 ATATTATTTTATTTATTTGCT 5476

RESULT 3
 AAD28369
 ID AAD28369 standard; DNA; 8011 BP.
 XX
 AC AAD28369;

22-APR-2002 (first entry)
 Human chemically treated genomic DNA #10.

Human; cytosolic; antidepressant; neuroleptic; noctropic; antiaddictive;
 adrenergic alpha-1C-receptor; cytosine methylation; therapy; alcoholism;
 behavioural disorder; neurological; psychiatric; cancer; schizophrenia;
 Tourette's syndrome; smoking; human immunodeficiency virus dementia;
 drug abuse; migraine; ds.

Homo sapiens.

WO200202809-A2.

10-JAN-2002.

02-JUL-2001; 2001WO-EP07540.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-154759/20.

Novel nucleic acid useful for diagnosis and therapy of behavioral
 disorder, neurological disorder and cancer, comprises a sequence of a
 PT segment of chemically pretreated DNA of adrenergic alpha-1C-receptor
 gene -

Claim 1; Page 60-64; 190bp; English.

The invention relates to nucleic acids comprising a segment of chemically
 CC pretreated DNA of adrenergic alpha-1C-receptor gene. The invention also
 CC relates to oligonucleotides or peptide nucleic acid (PNA) oligomers
 CC useful for detecting cytosine methylations. The pretreated DNA is useful
 CC for the diagnosis or therapy of behavioural disorders, neurological
 CC disorders and cancer, in particular major depressive disorder, Tourette's
 CC syndrome, schizophrenia, psychiatric and neurological disorders, smoking,
 CC drug abuse, alcoholism, personality traits, compulsive gambling, human
 CC immunodeficiency virus dementia, migraine, behaviours in schizophrenic
 CC and schizoaffective patients, and suicidal behaviour in patients with
 CC schizophrenia. The nucleic acid is useful for detecting the methylation
 CC state of all CpG dinucleotides and/or single nucleotide polymorphisms
 CC (SNPs). The present sequence is human chemically treated genomic DNA.

Sequence 8011 BP; 2439 A; 145 C; 1810 G; 3617 T; 0 other;

Query Match 1.2%; Score 21; DB 24; Length 8011;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTTTATTTATTTGCT 1379
 DB 5456 ATATTATTTTATTTATTTGCT 5476

RESULT 4
 ABL31320
 ID ABL31320 standard; DNA; 11907 BP.
 XX
 AC ABL31320;
 XX

DT 23-APR-2002 (first entry)
 XX
 DE Signal transduction associated gene modified DNA #82.
 XX
 KM Human, signal transduction associated gene; cytosine methylation state;
 KM CpG island; signal transduction associated disease; solid tumour; cancer;
 KM antitumour; cytostatic; mutant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 FN WO200200926-A2.
 PD
 PD 03-JAN-2002.
 PD
 PD 29-JUN-2001; 2001WO-EP07472.
 PD
 PD 30-JUN-2000; 2000DE-1032529.
 PD 01-SEP-2000; 2000DE-1043826.
 PD (EPIC-) EPIDNOMICS AG.
 PD
 PD Olek A, Piepenbrock C, Berlin K;
 PD WPI; 2002-147896/19.
 DR
 XX
 PT Oligonucleotide for diagnosis and therapy of diseases associated with
 PT signal transduction e.g. cancer, comprises chemically modified genomic
 PT sequences of genes associated with signal transduction
 PS
 PS Claim 1; SEQ ID No 163; 24pp; English.
 XX
 CC The present invention relates to chemically modified DNA sequences of
 CC signal transduction associated genes. The DNA sequences are chemically
 CC modified using a solution of bisulphite, hydrogen sulphite or
 CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers
 CC for detecting the cytosine methylation state (CpG islands) of these
 CC genes, and a method for the diagnosis and/or therapy of genetic and
 CC epigenetic parameters of genes associated with signal transduction.
 CC The genomic DNA can be obtained from cells or cellular components which
 CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,
 CC cerebrospinal fluid, tissue embedded in paraffin such as tissue from
 CC eye, intestine, kidney, brain, heart, prostate, lung, breast or liver,
 CC histologic object slides, and all their possible combinations. The
 CC sequences of the invention are useful for the diagnosis and therapy of
 CC diseases associated with signal transduction e.g. solid tumours and
 CC cancer. ABX31158-ABX3145 represent chemically pretreated genomic DNA
 CC sequences of different genes associated with signal transduction, or
 CC their complementary sequences.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 11907 BP; 3442 A; 180 C; 2430 G; 5855 T; 0 other;
 Query Match 1.2%; Score 21; DB 24; Length 11907;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1359 ATATTATTATATTTGCT 1379
 DB 2943 ATATTATTATATTTGCT 2963
 RESULT 5
 AAT58840
 ID AAT58840 standard; DNA; 580073 BP.
 XX
 AC AAT58840;
 XX
 DT 27-MAR-1997 (first entry)
 XX
 DE Mycoplasma genitalium genome.

XX M. genitalium; DNA; DNA gyrase; origin of replication; ORF; ss.
 KM megabase shotgun sequencing method; open reading frame; ss.
 XX
 XX Mycoplasma genitalium.
 PH
 PH Key
 CDS
 Location/Qualifiers
 8552..9184
 /tag= a
 /label= MG006
 /note= "Previously identified as MORF-20076, the
 encoded protein shows 27.59 percentage
 identity to thymidylate kinase (CDC8)
 from Saccharomyces cerevisiae"
 11252..12040
 /tag= b
 /label= MG009
 /note= "Previously identified as MORF-20078, the
 encoded protein shows 35.43 percentage
 identity to the Bacillus subtilis hypothetical
 protein covered in accession number
 G81D26185_102"
 12069..12725
 /tag= c
 /label= MG010
 /note= "Previously identified as MORF-20079, the
 encoded protein shows 25.73 percentage
 identity to DNA primase (dnaB) from
 Clostridium acetobutylicum"
 complement (13570..14247)
 /tag= d
 /label= MG012
 /note= "Previously identified as MORF-20080, the
 encoded protein shows 31.50 percentage
 identity to the ribosomal protein S6
 modification protein (rimK) from Escherichia
 coli"
 complement (14396..15217)
 /tag= e
 /label= MG013
 /note= "Previously identified as MORF-19823, MORF-20080
 and MORF-20081, the encoded protein shows 33.04
 percentage identity to 5,10-methylene-tetra-
 hydrofolate dehydrogenase (FOLD) from E. coli"
 17474..19243
 /tag= f
 /label= MG015
 /note= "Previously identified as MORF-20084, the
 encoded protein shows 32.23 percentage
 identity to transport ATP-binding protein
 (msbA) from E. coli"
 26478..27344
 /tag= g
 /label= MG023
 /note= "Previously identified as MORF-20092, the
 encoded protein shows 45.96 percentage
 identity to fructose-bisphosphate aldolase
 (tst) from B. subtilis"
 27345..28448
 /tag= h
 /label= MG024
 /note= "Previously identified as MORF-19826 and
 MORF-20093, the encoded protein shows 46.84
 percentage identity to GTP-binding protein
 from E. coli"
 36987..38978
 /tag= i
 /label= MG032
 /note= "Previously identified as MORF-20099, the
 encoded protein shows 26.82 percentage
 identity to ATP-dependent nuclease (addA)
 from B. subtilis"
 39242..39904
 CDS

```

FT      /*tag= j
FT      /label= MG033
FT      /note= "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (gipf) from B. subtilis"
FT      CDS
FT      complement (39873..40514)
FT      /tag= x
FT      /label= MG034
FT      /note= "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (tdk)
FT      from B. subtilis"
FT      CDS
FT      40543..41787
FT      /tag= l
FT      /label= MG035
FT      /note= "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hiss)
FT      from Mycobacterium leprae"
FT      CDS
FT      complement (44751..46277)
FT      /tag= m
FT      /label= MG038
FT      /note= "Previously identified as MORF-20105, the
FT      encoded protein shows 46.83 percentage
FT      identity to glycerol kinase (gipk)
FT      from E. coli"
FT      CDS
FT      complement (46268..47422)
FT      /tag= n
FT      /label= MG039
FT      /note= "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (GDH2) from S. cerevisiae"
FT      CDS
FT      49377..49643
FT      /tag= o
FT      /label= MG041
FT      /note= "The encoded protein shows 48.86 percentage
FT      identity to phosphohistidinoprotein-hexose
FT      phosphotransferase (prrh) from Mycoplasma
FT      capricolum"
FT      CDS
FT      50660..51520
FT      /tag= p
FT      /label= MG042
FT      /note= "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      CDS
FT      51525..52382
FT      /tag= q
FT      /label= MG043
FT      /note= "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (potB) from E. coli"
FT      CDS
FT      52366..53220
FT      /tag= r
FT      /label= MG044
FT      /note= "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein C (potC) from E. coli"
FT      CDS
FT      54658..55605
FT      /tag= s
FT      /label= MG046
FT      /note= "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to sialoglycoprotease (gcp)
FT      from Pasteurella haemolytica"
FT      CDS
FT      complement (56970..58310)
FT      /tag= t
FT      /label= MG048

```

```

FT      /note= "Previously identified as MORF-19834,
FT      MORF-20114 and MORF-20115, the encoded protein
FT      shows 43.02 percentage identity to signal
FT      recognition particle protein (fth) from B.
FT      subtilis"
FT      CDS
FT      58117..53079
FT      /tag= u
FT      /label= MG049
FT      /note= "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      CDS
FT      59083..53754
FT      /tag= v
FT      /label= MG050
FT      /note= "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deac) from Mycoplasma pneumoniae"
FT      CDS
FT      complement (64898..65731)
FT      /tag= w
FT      /label= MG056
FT      /note= "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185_99 from B. subtilis"
FT      CDS
FT      complement (65713..66249)
FT      /tag= x
FT      /label= MG057
FT      /note= "Previously identified as MORF-20123, the
FT      encoded protein shows 38.90 percentage
FT      identity to the protein disclosed in
FT      GB:D26185_104 from B. subtilis"
FT      CDS
FT      81047..82597
FT      /tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (SPase) from Staphylococcus aureus"
FT      CDS
FT      91065..91919
FT      /tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"
FT      Query Match
FT      1.2%; Score 21; DB 18; Length 580073;
FT      Best Local Similarity 100.0%; Pred. No. 13;
FT      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT      QY 1472 AACATGCTTTAACTTGT 1492
FT      DB 494176 AACATTGTTTAACTTGT 494196

```

RESULT 6

ABX09785
ID ABX09785 standard; DNA; 994 BP.

XX AC ABX09785;

XX DT 22-JAN-2003 (first entry)

XX DB M. incognita RNAi molecule #42 useful for nematode control.

XX KM RNAi molecule; double-stranded interfering RNA; nematode control;

XX KM RNA mediated interference; mRNA transcript; nematode gene;

XX KM growth; development; parasitism; reproduction; RNAi vector;

XX KM mRNA translation; nematode inhibitor; agricultural industry;

XX anti-nematode; ds.

OS Meiodiogyne incognita.
 XX WO200196584-A2.
 XX 20-DEC-2001.
 PD 12-JUN-2001; 2001WO-US18911.
 XX 12-JUN-2001; 2001WO-US18911.
 XX 12-JUN-2000; 2000US-210917P.
 PA (AKKA-) AKKADIX CORP.
 PI Mushaghi AR, Taylor CG, Feltelson JS, Eroshkin AM;
 DR WPI; 2002-139714/18.
 XX RNA mediated interference molecule useful for disrupting cellular
 PT process in a nematode, for controlling nematodes comprises genetic
 PT regulatory sequences -
 XX
 PS Claim 1; Page 96; 103pp; English.
 XX The present invention relates to RNAi (double-stranded interfering
 CC RNA or RNA mediated interference) molecules (nematode genes), and
 CC methods of using these sequences in nematode control. RNAi molecules
 CC selectively target mRNA transcripts of essential nematode genes. The
 CC RNAi molecules of the invention are useful for disrupting cellular
 CC processes in a nematode by contacting the nematode with a composition
 CC comprising an RNAi molecule. The RNAi molecules are useful for killing
 CC nematodes and/or inhibiting their growth, development, parasitism or
 CC reproduction and also for the regulation of levels of specific RNA
 CC in nematodes. Genetic regulatory sequences such as promoters, enhancers
 CC and terminators can be used in genetic constructs such as RNAi vectors
 CC which can be used for nematode control. The RNAi molecules are
 CC capable of targeting and reducing (and, in some cases, preventing)
 CC the translation of a specific gene product, and can be used to reduce
 CC or prevent mRNA translation in any tissue of the nematode because of
 CC its ability to cross tissue and cellular boundaries. The RNAi molecule
 CC can be contacted with a nematode by soaking, injection, or consumption
 CC of a food source containing an RNAi molecule. The RNAi molecules can
 CC also be used as an epigenetic factor to prevent the proliferation of
 CC subsequent generations of nematodes, to produce nematode inhibitors
 CC or RNAi in the plants, and provide new biotechnological strategies for
 CC managing nematodes under sustainable agricultural conditions.
 CC ABX09677-ABX09815 represent RNAi molecules useful for the control of
 CC nematodes.
 CC
 SQ Sequence 994 BP; 357 A; 155 C; 180 G; 302 T; 0 other;
 Query Match 1.1%; Score 20; DB 24; Length 994;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 943 TAATCATCATTTAATTA 962
 DB 21 TAATCATCATTTAATTA 40
 RESULT 7
 AAS63211/c
 ID AAS63211 standard; cDNA; 1738 BP.
 XX
 AC AAS63211;
 XX
 XX 29-JAN-2002 (first entry)
 DT
 XX Human purified secretory polynucleotide #67.
 DE
 XX Human; purified secretory polypeptide; cell proliferative disorder; ss;
 KM cancer; immune system disorder; neurological disorder; mental disorder;
 KM motor neuron disorder; demyelinating disorder; neuromuscular disorder;
 KM central nervous system disorder; enzyme linked immunosorbent assay;
 KM ELISA; gene therapy.

XX Homo sapiens.
 OS
 XX WO200162918-A2.
 XX
 XX 30-AUG-2001.
 PD
 XX
 XX 01-FEB-2001; 2001WO-US03465.
 PF
 XX 24-FEB-2000; 2000US-185215P.
 XX 24-FEB-2000; 2000US-185216P.
 PR 16-MAY-2000; 2000US-205232P.
 PR 17-MAY-2000; 2000US-205286P.
 PR 17-MAY-2000; 2000US-205287P.
 PR 17-MAY-2000; 2000US-205323P.
 PR 17-MAY-2000; 2000US-205324P.
 XX
 XX (INCYTE) INCYTE GENOMICS INC.
 PA
 XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 PI Chen A, D'sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE;
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockdreher TK, Daffo A;
 PI Wright RJ, Yap PE, Yu YJ, Bradley DL, Bratcher SR, Chen W;
 PI Cohen HD, Hodgson DW, Lincoln SE;
 DR WPI; 2001-648217/74.
 XX
 PT Nucleic acids encoding secretory polypeptides, useful in genetic
 PT diagnosis and therapy -
 XX
 PS Claim 1; Page 181-182; 237pp; English.
 XX Sequences AAS63145-AAS63223 represent DNA encoding purified secretory
 CC polypeptides of the invention. The polypeptides and polynucleotides can
 CC be used in the treatment, prevention and diagnosis of diseases associated
 CC with inappropriate secretory protein expression. These diseases include
 CC cell proliferative disorders such as atherosclerosis and psoriasis,
 CC cancers such as leukemia and melanoma, immune system disorders such as
 CC asthma and diabetes mellitus, neurological disorders such as epilepsy and
 CC parkinson's disease, mental disorders such as schizophrenia and seasonal
 CC affective disorder (SAD), motor neuron disorders such as amyotrophic
 CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,
 CC central nervous system disorders such as mental retardation and
 CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and
 CC muscular dystrophy. Target polynucleotides in a sample can be detected by
 CC hybridizing the sample with a probe sequence complementary to the target
 CC polynucleotide, under conditions in which a hybridisation complex is
 CC formed, and detecting the presence or absence of the complex. The
 CC polypeptides may also be used as antigens in the production of antibodies
 CC against secretory proteins and in assays to identify modulators of
 CC protein expression and activity. The antibodies may also be used as
 CC diagnostic agents for detecting the presence of the sequences of the
 CC invention in samples e.g. by enzyme linked immunosorbent assay (ELISA).
 CC
 SQ Sequence 1738 BP; 559 A; 357 C; 332 G; 490 T; 0 other;
 Query Match 1.1%; Score 20; DB 22; Length 1738;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1343 TCTCTCAATATATCGATAT 1362
 DB 620 TCTCTCAATATATCGATAT 601
 RESULT 8
 AAS51332/c
 ID AAS51332 standard; cDNA; 1741 BP.
 XX
 AC AAS51332;
 XX
 XX 21-OCT-2002 (first entry)
 DT

XX DE cDNA encoding human secretory protein #30.

XX KW Human; secretory polypeptide; SPM; actinic keratosis; arteriosclerosis;

XX KW buritis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;

XX KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;

XX KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;

XX KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;

XX KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;

XX KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;

XX KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;

XX KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;

XX KW Parkinson's disease; central nervous system disorder; mental disorder;

XX KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;

XX KW gene therapy; gene; ss.

XX OS Homo sapiens.

XX PN WO200257304-A2.

XX PD 25-JUL-2002.

XX PF 15-JAN-2002; 2002WO-US01340.

XX PR 16-JAN-2001; 2001US-261864P.

XX PR 16-JAN-2001; 2001US-261865P.

XX PR 16-JAN-2001; 2001US-261979P.

XX PR 16-JAN-2001; 2001US-261981P.

XX PR 17-JAN-2001; 2001US-262164P.

XX PR 17-JAN-2001; 2001US-262208P.

XX PR 17-JAN-2001; 2001US-263131P.

XX PR 19-JAN-2001; 2001US-262598P.

XX PR 19-JAN-2001; 2001US-262780P.

XX PR 19-JAN-2001; 2001US-263063P.

XX PR 19-JAN-2001; 2001US-263066P.

XX PR 19-JAN-2001; 2001US-263069P.

XX PR 19-JAN-2001; 2001US-263070P.

XX PR 19-JAN-2001; 2001US-263074P.

XX PR 19-JAN-2001; 2001US-263076P.

XX PR 19-JAN-2001; 2001US-263077P.

XX PR 19-JAN-2001; 2001US-263329P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;

XX PI Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwana R, Chen AJ;

XX PI Chang SC, Gerstein EH, Peralta CH, David MR, Lewis SA;

XX DR WPI; 2002-590716/63.

XX DR N-PSDB; ABG69840, ABG69841.

XX PT New purified secretory polypeptides and polynucleotides, useful in the

XX PT diagnosis, study, prevention or treatment of diseases associated with

XX PT decreased expression of functional secretory molecules, e.g. AIDS,

XX PT cancer or allergies

XX PS Claim 1; Page 266; 340P; English.

XX CC The invention describes an isolated polynucleotide a naturally occurring

XX CC polynucleotide sequence at least 90 % identical to it, a polynucleotide

XX CC complementary to it or an RNA equivalent of it. The purified secretory

XX CC polypeptides (SPM) and polynucleotides are useful in the diagnosis,

XX CC study, prevention or treatment of diseases associated with decreased

XX CC expression of functional SPM, e.g. actinic keratosis, arteriosclerosis,

XX CC buritis, cirrhosis, hepatitis, polycythaemia vera, primary

XX CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,

XX CC leukaemia, myeloma or sarcoma, immune system disorder such as acquired

XX CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,

XX CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,

XX CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,

XX CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid

XX CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's

XX CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's

CC disease, other developmental disorder of the central nervous system,

CC mental disorder including mood, anxiety or schizophrenic disorder,

CC amnesia or Tourette's disorder. The polynucleotides may be used in

CC hybridisation and amplification technologies, e.g. in assessing gene

CC expression patterns, to develop a transcript image for a particular cell

CC or tissue, or to create transgenic animals to model human disease. This

CC sequence encodes a human secretory protein isolated in the invention.

XX SQ Sequence 1741 BP; 560 A; 357 C; 333 G; 491 T; 0 other;

XX Query Match 1.1%; Score 20; DB 24; Length 1741;

XX Best Local Similarity 100.0%; Pred. No. 49;

XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Db 1343 TCTCTCAATATATCTGATAT 1362

XX 622 TCTCTCAATATATCTGATAT 603

XX RESULT 9

XX ID ABZ17008/c

XX ID ABZ17008 standard; DNA; 2000 BP.

XX AC ABZ17008;

XX DT 21-JAN-2003 (first entry)

XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4813.

XX DE Arabidopsis thaliana, plant; gene; stress; transgenic; ds.

XX OS Arabidopsis thaliana.

XX PN WO200216655-A2.

XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26685.

XX PR 24-AUG-2000; 2000US-227866P.

XX PR 26-JAN-2001; 2001US-264647P.

XX PR 22-JUN-2001; 2001US-300111P.

XX PA (SCRI) SCRIPPS RES. INST.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Harper JF, Krops J, Wang X, Zhu T;

XX DR WPI; 2002-304127/34.

XX PT Identifying a stress condition to which a plant cell has been exposed

XX PT and producing plants with increased tolerance to these abiotic stresses

XX PS Claim 144; SEQ ID NO 4813; 577bp + Sequence Listing; English.

XX CC The invention relates to identifying a stress condition to which a plant

XX CC cell has been exposed, comprising:

XX CC (a) contacting nucleic acid representative of expressed polynucleotides

XX CC in the plant cell with an array or probes representative of the plant

XX CC cell genome; and

XX CC (b) detecting a profile of expressed polynucleotides in the plant cell

XX CC characteristic of a stress response. The method is useful in the

XX CC production of transgenic plants, cells and seeds and in producing plants

XX CC with increased tolerance to abiotic stress. The present sequence is that

XX CC of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used

XX CC in methods of the invention.

XX CC Note: The sequence data for this patent is not represented in the printed

XX CC specification but is based on sequence information supplied to Derwent by

XX CC the European Patent Office.

XX SQ Sequence 2000 BP; 723 A; 323 C; 280 G; 674 T; 0 other;

Query Match 1.1%; Score 20; DB 24; Length 2000;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 AAAGATTATTAAGTGAAT 238
 |||||
 DB 558 AAAGATTATTAAGTGAAT 539

RESULT 10
 AAD29001
 ID AAD29001 standard; DNA; 5686 BP.
 XX
 AC AAD29001;
 XX
 DT 07-MAY-2002 (first entry)
 XX
 DE Bugula neritina contig 5 DNA sequence from cosmid 6A.
 XX
 KW Polyketide; bryopyran ring; byrostatin; breast cancer; anticancer;
 KW antifungal; antimicrobial; immunomodulatory; polyketide synthase;
 KW PKS; enzyme; ds.
 XX
 OS Bugula neritina.
 XX
 PN WO200111024-A2.
 XX
 PD 15-FEB-2001.
 XX
 PF 04-AUG-2000; 2000WO-US21326.
 XX
 PR 04-AUG-1999; 99US-147283P.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 PI Haygood M, Davidson SK, Allen SW, Hildebrand M;
 XX
 DR WPI; 2002-154285/20.
 XX
 PT Composition comprising a polypeptide isolated from marine organism,
 PT which catalyses at least one step in synthesis of polyketide/bryopyran
 PT ring, useful for producing polyketide or bryopyran ring containing
 PT compositions -
 XX
 PS Claim 64; Fig 15B; 233p; English.
 XX
 CC The invention relates to compositions comprising nucleic acid molecules
 CC encoding a polypeptide which catalyses at least one step in synthesis of
 CC polyketides including bryopyran ring, such as byrostatis. These novel
 CC sequences are derived from marine organisms. Compositions containing
 CC sequences of the invention are useful for producing base structure,
 CC bryopyran rings that can form the basis of combinatorial chemistry to
 CC form a wide variety of compounds which can be screened for bioactivities
 CC including anticancer activity. The cloned genes and linked genes
 CC involved in byrostatin synthesis can be used to screen environmental
 CC samples for polyketide synthase (PKS) genes. They are also used for
 CC combinatorial creation of novel polyketide/byrostatin analogues that may
 CC exhibit improved anti-cancer properties. Compositions of the invention
 CC are useful for producing byrostatin and its analogues which are useful
 CC for treating breast cancer and as anticancer, antifungal antimicrobial
 CC and immunomodulatory compounds. They are useful for producing novel
 CC polyketides such as bryopyran rings including byrostatis. The present
 CC sequence is Bugula neritina contig 5 DNA sequence from cosmid 6A
 CC used in the exemplification of the invention.
 CC
 SO Sequence 5686 BP; 1601 A; 924 C; 1195 G; 1514 T; 452 other;
 XX
 QY
 Query Match 1.1%; Score 20; DB 24; Length 5686;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1052 TATGTTACTTTCTTCAAGC 1071

DB 4061 TATGTTACTTTCTTCAAGC 4060
 RESULT 11
 ABL32028
 ID ABL32028 standard; DNA; 7351 BP.
 XX
 AC ABL32028;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Human immune system associated gene SEQ ID NO: 1.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antiidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200200928-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -
 XX
 PS Claim 1; SEQ ID NO 1; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 CC
 SO Sequence 7351 BP; 2230 A; 42 C; 1230 G; 3849 T; 0 other;
 XX
 QY
 Query Match 1.1%; Score 20; DB 24; Length 7351;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 969 TTATTTAAATATATGATATA 988
 |||||
 DB 6500 TTATTTAAATATATGATATA 6519

RESULT 12
 ABL40057/C
 ID ABL40057 standard; DNA; 7441 BP.
 XX
 AC ABL40057;
 XX
 DT 21-MAY-2002 (first entry)
 XX


```
DE Human chemically pretreated gene sequence #70 strand 1.
XX
XX
XX Human; de; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW cytoskeletal; ALDH6; CYP11A; CYP3A3; DYPD; EPHX2; OCLN; TXNRD1;
KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
XX Homo sapiens.
XX
XX WO200202806-A2.
XX
XX 10-JAN-2002.
XX
XX 29-JUN-2001; 2001WO-EP07470.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-154757/20.
XX
XX
XX New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT useful for detecting cytosine methylation state of genes associated
PT with pharmacogenomics and for therapy of diseases e.g. cancer -
XX
XX
XX Claim 1; SEQ ID No 139; 24pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence at
CC least 18 bases in length of a segment of the chemically pretreated DNA
CC of genes associated with pharmacogenomics according to one of the
CC sequences of the genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B1
CC (NM 000497), CYP3A3 (NM 000776 and NM 017460), DYPD (NM 000110), EPHX2
CC (NM 001979), OCLN (NM 002538), TXNRD1 (NM 003330), UGT8 (NM 003360)
CC MRP (NM 004996, NM 015900, NM 019901, NM 019902, NM 019862, NM 019898,
CC NM 019899) and their complementary sequences, or a sequence (SI) chosen
CC from 87 sequences and their complements. The chemical pretreatment
CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
CC into uracils. Also included are an oligomer (II) in particular an
CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
CC each case at least one base sequence having a length of 9 nucleotides
CC which hybridises to or is identical to a chemically pretreated DNA of
CC genes associated with pharmacogenomics and their complements, arranged in
CC an array for analysing diseases associated with the methylation state
CC (CpG) and/or detecting SNPs (single nucleotide polymorphisms)
CC of the 87 sequences. The oligomers may also be used as PCR primers.
CC The set of 87 nucleic acids and their complements is useful for diagnosis
CC and therapy of solid tumours and cancer. The present sequence
CC represents one the 87 DNA sequences or its complement.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 7441 BP; 2098 A; 66 C; 1948 G; 3329 T; 0 other;
SQ
Query Match 1.1%; Score 20; DB 24; Length 7441;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1218 TTTACACAAATATATAATA 1237
Db 6788 TTTACACAAATATATAATA 6769
RESULT 13
AAS63344/c
ID AAS63344 standard; DNA; 7479 BP.
XX
AC AAS63344;
XX
XX 29-JAN-2002 (first entry)
XX
XX Chemically pretreated metabolism associated gene #39.
DE
XX
XX Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;
KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;
KW single nucleotide polymorphism detection; SNP; stool; urine; lung;
KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;
KW DUSP2; EPHX2; QDPR; SGGH; SHMT2; SLC7A2; SLC7A4; TYMS; de.
XX
XX Homo sapiens.
XX
XX WO200176451-A2.
XX
XX 18-OCT-2001.
XX
XX 06-APR-2001; 2001WO-EP04016.
XX
XX 06-APR-2000; 2000DE-1019058.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-010834/01.
XX
XX
XX New nucleic acid, useful for diagnosis and therapy of metabolic
PT disease, solid tumour and cancers, comprises segment of chemically
PT modified genomic sequences of genes associated with metabolism -
XX
XX
XX Claim 1; Page 107-109; 143pp; English.
XX
XX The invention relates to a nucleic acid (II) comprising a sequence at
CC least 18 bases of a segment of the chemically pretreated DNA of genes
CC associated with metabolism such as DUSP2 (NM 004418), EPHX2 (NM 001979),
CC QDPR (NM 000320), SGGH (NM 000199), SHMT2 (NM 005412), SLC7A2
CC (NM 003046), SLC7A4 (NM 004173) and TYMS (NM 001071) (all
CC undefined). (I) are useful for diagnosis and therapy of metabolic
CC disease, solid tumours and cancers; as primer oligonucleotides for the
CC amplification of DNA sequences, for detecting the cytosine methylation
CC state and/or single nucleotide polymorphisms (SNPs) in a chemically
CC treated DNA of genes associated with metabolism. An array of (I) is
CC useful for ascertaining genetic and/or epigenetic parameters for the
CC diagnosis and/or therapy of existing diseases or the predisposition to
CC specific diseases by analysing cytosine methylations. The method involves
CC chemically treating genomic DNA sample by a solution of bisulphite,
CC hydrogen sulphite or disulphite such that cytosine bases which are
CC unmethylated at the 5th-position are converted to uracil or another base
CC which is dissimilar to cytosine in terms of hybridisation behaviour and
CC amplifying fragments of the chemically pretreated genomic DNA. The
CC genomic DNA is from cells or cellular components which contain DNA,
CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,
CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as
CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast
CC or liver, histologic object slides and their combinations. Genetic
CC parameters are mutations, in particular insertions, deletions, point
CC mutations, inversions and polymorphisms of genes associated with
CC metabolism and sequences further required for their regulation.
CC Epigenetic parameters are in particular cytosine methylations and
CC further chemical modifications of DNA bases of genes associated with
CC metabolism. Further epigenetic parameters include for e.g. the
CC acetylation of histones which correlates with DNA methylation.
CC AAS63306-AAS63373 represent chemically pretreated metabolism associated
CC genes, and related primers of the invention.
XX
XX Sequence 7479 BP; 2111 A; 68 C; 1962 G; 3338 T; 0 other;
SQ
Query Match 1.1%; Score 20; DB 24; Length 7479;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1218 TTTACACATATATAATA 1237
 |||||
 DB 6826 TTTACACATATATAATA 6807

RESULT 14
 ID ABRN0086 standard; DNA; 7922 BP.
 AC ABRN0086;
 XX
 DT 15-JUL-2002 (first entry)
 XX
 DE Human chemically modified disease associated gene SEQ ID NO 103.
 XX
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KW antidiabetic; cyrostatic; anticonvulsant; ds.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200200927-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07536.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130908/17.
 XX
 PT Novel nucleic acid useful for diagnosis and therapy of diseases
 PT associated with development genes such as diabetes, comprises a
 PT sequence of a segment of chemically pretreated DNA of genes associated
 PT with development
 XX
 PS Claim 1; SEQ ID NO 103; 27bp; English.
 XX
 CC The invention relates to a nucleic acid (i) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (ii)
 CC of genes associated with development selected from 87 genes listed in
 CC the specification such as ACCFN, ADPN, or AFDI and comprising one of 350
 CC sequences (ABN79984-ABN80333) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular disease related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Currarino syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or
 CC (ii) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (ii) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but is based on sequence information supplied to Derwent by
 CC the European Patent Office.
 XX
 SQ Sequence 7922 BP; 2094 A; 210 C; 1947 G; 3670 T; 1 other;

Query Match 1.1%; Score 20; DB 24; Length 7922;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 918 TTTAGTAGTATTATAGAA 937

DB 1568 TTTAGTAGTATTATAGAA 1587
 |||||

RESULT 15
 ID ABL29482 standard; DNA; 32548 BP.
 AC ABL29482;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 39919.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX
 PS Claim 1; SEQ ID NO 39919; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
 CC sequences (AB101840-AB116175) and the encoded proteins
 CC (AB557737-AB12072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WPI
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 32548 BP; 9803 A; 6479 C; 6592 G; 9674 T; 0 other;

Query Match 1.1%; Score 20; DB 23; Length 32548;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1108 GAAAAAATGTATATGAAT 1127
 |||||
 DB 30680 GAAAAAATGTATATGAAT 30699

Search completed: November 25, 2003, 17:02:58
 Job time : 370 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 15:19:58 / Search time 4396 Seconds

(without alignments)
16592.761 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccactctctgactcttga.....tagagcgaacgaactgaacc 1783

Scoring table: OLIGO_NTC
Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rdc:*
36: em_hlg_mam:*
37: em_hlg_vrt:*
38: em_sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 24 | 1.3 | 180097 | 9 | AC1233978 |
| 2 | 24 | 1.3 | 227691 | 2 | AC110885 |
| 3 | 23 | 1.3 | 46067 | 10 | AF539527 |
| 4 | 23 | 1.3 | 120648 | 8 | AP003859 |
| 5 | 23 | 1.3 | 130689 | 2 | AL806534 |
| 6 | 23 | 1.3 | 169888 | 2 | AC115848 |
| 7 | 23 | 1.3 | 303446 | 1 | AB016941 |
| 8 | 22 | 1.2 | 2282 | 8 | VNA277287 |
| 9 | 22 | 1.2 | 67271 | 2 | AC133887 |
| 10 | 22 | 1.2 | 108653 | 9 | HSDJ863C7 |
| 11 | 22 | 1.2 | 153537 | 2 | AC121962 |
| 12 | 22 | 1.2 | 160536 | 2 | AC040953 |
| 13 | 22 | 1.2 | 165870 | 9 | AC025089 |
| 14 | 22 | 1.2 | 175667 | 9 | AC009336 |
| 15 | 22 | 1.2 | 194149 | 9 | AC100799 |
| 16 | 22 | 1.2 | 201923 | 10 | AC121772 |
| 17 | 22 | 1.2 | 201935 | 10 | AC121582 |
| 18 | 22 | 1.2 | 206975 | 2 | AC133827 |
| 19 | 22 | 1.2 | 212024 | 2 | AC073390 |
| 20 | 22 | 1.2 | 224846 | 10 | AC125327 |
| 21 | 22 | 1.2 | 236579 | 2 | AC121707 |
| 22 | 22 | 1.2 | 245226 | 2 | AC103003 |
| 23 | 22 | 1.2 | 290670 | 2 | AC121043 |
| 24 | 21 | 1.2 | 8011 | 6 | AX344981 |
| 25 | 21 | 1.2 | 8011 | 6 | AX344981 |
| 26 | 21 | 1.2 | 9060 | 1 | U39721 |
| 27 | 21 | 1.2 | 11907 | 6 | AX344316 |
| 28 | 21 | 1.2 | 19366 | 7 | PZACG |
| 29 | 21 | 1.2 | 36087 | 3 | L16685 |
| 30 | 21 | 1.2 | 42732 | 9 | AC007206 |
| 31 | 21 | 1.2 | 43157 | 9 | AC079466 |
| 32 | 21 | 1.2 | 80059 | 9 | AC114940 |
| 33 | 21 | 1.2 | 90500 | 9 | AC069235 |
| 34 | 21 | 1.2 | 100226 | 9 | AL583802 |
| 35 | 21 | 1.2 | 108537 | 2 | AC139355 |
| 36 | 21 | 1.2 | 110000 | 3 | AC116957_0 |
| 37 | 21 | 1.2 | 110000 | 6 | AC116957_1 |
| 38 | 21 | 1.2 | 110000 | 6 | AR300198_4 |
| 39 | 21 | 1.2 | 125400 | 9 | AC1233904 |
| 40 | 21 | 1.2 | 129451 | 9 | AL359474 |
| 41 | 21 | 1.2 | 134804 | 9 | AC004879 |
| 42 | 21 | 1.2 | 137783 | 9 | AL441884 |
| 43 | 21 | 1.2 | 139378 | 9 | HS64K7 |
| 44 | 21 | 1.2 | 142805 | 9 | AL592525 |
| 45 | 21 | 1.2 | 147038 | 9 | AL133326 |

ALIGNMENTS

RESULT 1
AC1233978
LOCUS
AC1233978
DEFINITION
Papio anubis clone rp41-211d5, complete sequence.
ACCESSION
AC1233978
VERSION
AC1233978.5
GI:22038600
KEYWORDS
HTG.
SOURCE
Papio anubis (olive baboon)
ORGANISM
Papio anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopitheciinae; Papio.
REFERENCE
1 (bases 1 to 180097)
Song, L. and Roe, B.A.
Papio anubis BAC Clone rp41-211d5

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 180097)
 AUTHORS Song, L. and Roe, B. A.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUN-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA

REFERENCE 3 (bases 1 to 180097)
 AUTHORS Song, L. and Roe, B. A.
 TITLE Direct Submission
 JOURNAL Submitted (19-JUL-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA

REFERENCE 4 (bases 1 to 180097)
 AUTHORS Song, L. and Roe, B. A.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
 OK 73019, USA

COMMENT On Aug 1, 2002 this sequence version replaced gi:21909512.
 ----- Genome Center
 Center: Department Of Chemistry And Biochemistry
 The University Of Oklahoma
 Center code: UOKNOR

FEATURES
 Source
 1..180097
 /organism="Papio anubis"
 /mol_type="genomic DNA"
 /sub_species="anubis"
 /db_xref="taxon:9555"
 /clone="rp41-211d5"
 /clone_lib="RPC1 - 41 Male (Olive) Baboon BAC Library"

BASE COUNT 48904 a 39123 c 39387 g 52683 t

ORIGIN

Query Match 1.3%; Score 24; DB 9; Length 180097;
 Best Local Similarity 100.0%; Pred.No. 1.2;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 314 ATTCTTACACGTGAGAAATTT 397
 |||||
 Db 115577 ATTCTTACACGTGAGAAATTT 115600
 |||||

RESULT 2
 LOCUS AC110885 227691 bp DNA linear HTG 21-AUG-2002
 DEFINITION Mus musculus clone RP24-342P18, WORKING DRAFT SEQUENCE, 9 ordered
 pieces.
 AC110885
 AC110885.5 GI:22381716
 HTG: HTGS PHASE3; HTGS DRAFT; HTGS_FULLTOP.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 227691)
 Birren, B., Nusbaum, C. and Lander, E.
 Mus musculus clone RP24-342P18
 Unpublished
 2 (bases 1 to 227691)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckigalter, B.,
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
 Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Farc, S.,
 Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S.,
 Girde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
 MacDonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M.,

McKernan, K., Meldrim, J., Meneus, L., Mihova, T.,
 Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
 Peterson, K., Phunkhang, P., Pierris, N., Pollara, V., Raymond, C.,
 Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,
 Severy, P., Spencer, B., Stange-Thomann, J., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talmas, J., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. C., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 TITLE
 JOURNAL Submitted (16-FEB-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 227691)
 Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckigalter, B.,
 Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Farc, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
 Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
 Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,
 McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
 Phunkhang, P., Pierris, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
 Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
 Smith, C., Spencer, B., Stange-Thomann, J., Stojanovic, N., Talmas, J.,
 Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
 Zembek, L., Zimmer, A. and Zody, M.
 DIRECT SUBMISSION
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 21, 2002 this sequence version replaced gi:20806336.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: 119395
 Center clone name: 342.P.18

----- Summary Statistics
 Sequencing vector: Plasmid; N/A; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 225012 bases at least Q40
 Consensus quality: 226175 bases at least Q30
 Consensus quality: 226632 bases at least Q20
 Insert size: 225000; agarose-fp
 Insert size: 226891; sum-of-contigs
 Quality coverage: 12.1 in Q20 bases; agarose-fp
 Quality coverage: 12.0 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 2355: contig of 2355 bp in length
 * 2356 2455: gap of 100 bp
 * 2456 10338: contig of 7883 bp in length
 * 10339 10438: gap of 100 bp

| | | | | | |
|-----------------------|--------------|--|----------------|--------------------|-----------------------|
| | * | 10439 | 17967: | contig of 7529 bp | in length |
| | * | 17968 | 18067: | gap of 100 bp | |
| | * | 18068 | 27410: | contig of 9343 bp | in length |
| | * | 27411 | 27510: | gap of 100 bp | |
| | * | 27511 | 44269: | contig of 16759 bp | in length |
| | * | 44270 | 44369: | gap of 100 bp | |
| | * | 44370 | 65178: | contig of 20809 bp | in length |
| | * | 65179 | 65279: | gap of 100 bp | |
| | * | 65279 | 89146: | contig of 23868 bp | in length |
| | * | 89147 | 89246: | gap of 100 bp | |
| | * | 89247 | 137077: | contig of 47831 bp | in length |
| | * | 137078 | 137177: | gap of 100 bp | |
| | * | 137178 | 227691: | contig of 90514 bp | in length. |
| FEATURES | | | | | |
| source | | | | | |
| | | 1..227691 | | | |
| | | /organism="Mus musculus" | | | |
| | | /mol_type="genomic DNA" | | | |
| | | /db_xref="taxon:10090" | | | |
| | | /clone="RP24-342P18" | | | |
| | | /clone_1lb="RPCI-24 Male Mouse BAC" | | | |
| | misc_feature | 1..2355 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 2456..10338 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 10439..117967 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 18068..27410 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 27511..44269 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 44370..65178 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 65279..89146 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 89247..137077 | | | |
| | | /note="assembly_fragment" | | | |
| | misc_feature | 137178..227691 | | | |
| | | /note="assembly_fragment" | | | |
| | | vector_end:r7 | | | |
| | | vector_side:right" | | | |
| BASE COUNT | | 61965 a | 50768 c | 49660 g | 64498 t 800 others |
| ORIGIN | | | | | |
| Query Match | | 1.3% | Score 24; | DB 2; | Length 227691; |
| Best Local Similarity | | 100.0%; | Pred. No. 1.1; | | |
| Matches | 24; | Conservative | % 0; | Mismatches | 0; Indels 0; Gaps 0; |
| Oy | 972 | TTTAAATTATGATGAAGTTGGAT | 995 | | |
| Dd | 161120 | TTTAAATTATGATGAAGTTGGAT | 161143 | | |
| RESULT 3 | | | | | |
| AF539527 | | | | | |
| LOCUS | | | | | |
| DEFINITION | | | | | |
| | | Mus musculus strain C57BL/6 O-linked GlcNAc transferase gene, | | | |
| | | complete cds. | | | |
| VERSION | | | | | |
| AFF539527 | | | | | |
| KEYWORDS | | | | | |
| AF539527.1 | | | | | |
| GI:27499605 | | | | | |
| SOURCE | | | | | |
| ORGANISM | | | | | |
| | | | | | |
| | | Mus musculus (house mouse) | | | |
| | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | | |
| | | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. | | | |
| | | 1 (bases 1 to 46067) | | | |
| | | Hanover,U.A., Yu,S., Lubas,W.B., Shin,S.H., Ragano-Cacciola,M., | | | |
| | | Kochian,J. and Love,D.C. | | | |
| | | Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc | | | |
| | | transferase encoded by a single mammalian gene | | | |
| | | Arch. Biochem. Biophys. 409 (2), 287-297 (2003) | | | |
| | | 12504895 | | | |
| | | (bases 1 to 46067) | | | |
| JOURNAL | | | | | |
| PIRMEED | | | | | |
| REFERENCE | | | | | |
| AUTHORS | | | | | |
| | | Hanover,U.A., Yu,S., Lubas,W.B., Shin,S.-H., Ragano-Cacciola,M., | | | |

TITLE J.K. and Love,D.C.
JOURNAL Direct Submission
 Submitted (09-AUG-2002) LCB, NIDDK, NIH, Bldg 8 Rm 402, 9000
 Rockville Pike, Bethesda, MD 20892-0851, USA
FEATURES Location/Qualifiers
 source
 1. .46067
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /chromosome="X"
 /map="XD"
mRNA
 join(11309, .1554,4215, .4395,5188, .5431,13392, .13460, 16877, .16993,22505, .22584,23623, .23818,24590, .24730, 25393, .25493,28017, .28170,28255, .28356,28719, .28898, 30500, .30658,30774, .30860,33191, .33316,33786, .33956, 34066, .34182,34285, .34455,36170, .36352,38732, .38984, 39247, .39370,41908, .44103)
 /product="O-linked G1CNAc transferase"
 join(1518, .1554,4215, .4395,5188, .5431,13392, .13460, 16877, .16993,22505, .22584,23623, .23818,24590, .24730, 25393, .25493,28017, .28170,28255, .28356,28719, .28898, 30500, .30658,30774, .30860,33191, .33316,33786, .33956, 34066, .34182,34285, .34455,36170, .36352,38732, .38984, 39247, .39370,41908, .42082)
 /codon_start=1
 /product="O-linked G1CNAc transferase"
 /protein_id="AA017363.1"
 /db_xref="GI:27495606"
 /translation="MASSGVNADSTPPTRMISFOGLAEIAREYQAGDEAAERHC
 MQLRQEPDNTGALLILTSIHFOCRRLDRSAHSTLAKONPLIAEYASNLGVYKER
 GQGEALIEHYRHALRLKPDYIDGYNIAALVAAGDEGVAOYASALQYNPDIQYRA
 SDGLNLEKALGRLEBAKAYCIKAETOPNFAVMSNLCYENAGCEIWLAIHFEEKAV
 LIDNPFDAIYNIGVLEKARIPDRAVAALRALSPNHAIVGNLACYAYEGGLID
 LAIDYRRALTELOHPFDPAQCNLANIKEGSVAAEDQYTLRLCPDASHNNLNA
 NIKRQENISEAYRLKALEVEPPEAAHNSLASVUQQGKLOEALMHKEARIESE
 TPDAAVSMKNTLKENQDVQALQCYTRAIQINPEFADASHNLASHKDSNLFEALA
 SYRALNLKPDPEPACNLACHQIVCDWTDYDERMKLVSIAEQLKRLPSPVHPH
 HSMYLLSHGFRKAIARHGNLCLDKLNLVLRKPYEPKDLKSDGLRAGVYSDSG
 NHFPMHMOGIPKEMNDKEVXCALSPDDGNFRKYVAEAAHFDLSQIPNGKRA
 ADRLHOGHILVNMNGTARGARELPAIPAOAWLIGPGSGALPMYITITDDE
 TSPREAYOYSEKILAMYPHPIGDMANPKKKAVIDPKSGHLYNRYVINGID
 LKAFDLSLPVKYVWKCPDGGNDPDSNTALMNPVTPMTTIAVETMINRGITQT
 INGSISNGATQIINNKAATGEVEPTIIIVTTSQYGLPEDAIVYCNFQLYKIDS
 TLQWANIILKRVPSVLMILRPFAVEGSPNTQYAAQNNGLPQNRILFSPVAKEXHVR
 GOLADVICLDIPLCNGHTGMDVMAGTPTMTGSETILASVASQLTCLCIELIAS
 ROEVEDIAVLKGLDLEYLKKIRGKVMQRISSPLEFNXXQYTMELERLYLQWMEHYAAG
 NKPHMFKPEVTESA"
BASE COUNT 12287 a 9079 c 9503 g 15155 t 43 others
ORIGIN
 Query Match 1.3%; Score 23; DB 10; Length 46067;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 430 AAACCTGACCTTTGACTTTCTTA 452
 |||||
 |||||
Db 22853 AAACCTGACCTTTGACTTTCTTA 22875
RESULT 4
LOCUS AP003859 120648 bp DNA linear PLN 12-JUN-2003
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,
 BAC clone:OU1033_B09, complete sequence.
ACCESSION AP003859
VERSION AP003859.2 GI:31621035
KEYWORDS HTG.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophytes; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Echinoideae; Oryzaceae; Oryza.

REFERENCE 1
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE *Oryza sativa nipponbare* (933) genomic DNA, chromosome 8, BAC
 clone:OJ1033_B09
 JOURNAL Published Only in Database (2001)
 REFERENCE 2 (bases 1 to 120648)
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (09-JUL-2001) Takuji Sasaki, National Institute of
 Agricultural Sciences, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rsgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT On Jun 12, 2003 this sequence version replaced gi:14646792.
 The orientation of the sequence is from M13rev to -21M13 of the BAC
 clone. The nucleotide sequence of this BAC clone was generated by
 combining Monsanto and RGP-Japan sequencing data.

FEATURES
 source
 1..120648
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="8"
 /clone="OJ1033_B09"

BASE COUNT 35274 a 25678 c 25266 g 34430 t
 ORIGIN

Query Match 1.3%; Score 23; DB 8; Length 120648;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1363 TTATTTATATTTGTGCATATT 1385
 Db 88378 TTATTTATATTTGTGCATATT 88356

RESULT 5
 LOCUS AL806534 130689 bp DNA linear ROD 19-SEP-2002
 DEFINITION Mouse DNA sequence from clone RP23-334A11 on chromosome X, complete
 sequence.
 ACCESSION AL806534 GI:23304104
 VERSION AL806534.3
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 130689)
 REFERENCE Blakey, S.
 AUTHORS Direct Submission
 JOURNAL Submitted (17-SEP-2002) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Sep 23, 2002 this sequence version replaced gi:22798318.

COMMENT ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em, EMBL; Sw,
 SWISSPROT; Tr, TREMBL; Wp, WORMPEB; Information on the WORMPEB
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpeb RP23-334A11 is
 from the RPI-23 Mouse PAC Library
 constructed by the group of Pieter de Jong.
 For further details see <http://www.choi.org/bacpac/home.htm>
 VECTOR: pBAC3.6

FEATURES
 source
 1..130689
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP23-334A11"
 /clone_11b="RPI-23"

BASE COUNT 37078 a 25556 c 27928 g 40127 t
 ORIGIN

Query Match 1.3%; Score 23; DB 10; Length 130689;
 Best Local Similarity 100.0%; Pred. No. 3.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 430 AAACCTTGACTTTGACTTCTTA 452
 Db 119793 AAACCTTGACTTTGACTTCTTA 119815

RESULT 6
 LOCUS AC115848 169888 bp DNA linear HTG 23-APR-2003
 DEFINITION Mus musculus clone RP24-248L23, WORKING DRAFT SEQUENCE, 12 ordered
 pieces.
 ACCESSION AC115848 GI:30017915
 VERSION AC115848.3
 KEYWORDS HTG, HTGS, PHAS2, HTGS, DRAFT.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 169888)
 REFERENCE Birren, B., Nusbaum, C. and Lander, E.
 AUTHORS Mus musculus, clone RP24-248L23
 TITLE Unpublished
 JOURNAL 2 (bases 1 to 169888)
 REFERENCE Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barin, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
 Charato, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
 Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kamat, A., Karatas, A., Kelle, C., Lacroque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., Macdonald, P., Major, J., Margulis, N., Matthews, C.,
 McCarthy, W., McEwan, P., McKernan, K., Meldrum, J., Mercus, L.,
 Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicot, R.,
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhan, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Theodore, J.,
 Strauss, K., Sudrmanian, A., Talamas, J., Tesfaye, S., Vasiliiev, H.,
 Topham, K., Travers, M., Travis, N., Triggilo, J., Vassiliiev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 TITLE Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA

gene
CDS
/locus_tag="BT3514"
178..3126
/locus_tag="BT3514"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA078620.1"
/db_xref="GI:29340830"
/translation="MRILPLVLVLTSLTALQSOVTVIKENALQSKAFPLVTSKAAV
VYDIDYLVKRTAKLEFSDVESVTGQQLRDLTKLGDKEIIIVGVENKRLIRQAE
KGLDLSLEGTEWERFLQTVSRPFGVSKALVAGSDRGAAYGAFSLSEMGVSPW
VMADVPYKGGKTLVYDAPATLSKTPSVKRGIFLNDMDGLKPMKATKFEKERGIG
PRVYAKICELLRLKANHLAPAMVPSTAFYKIPENKLVADTFAIAGSSHOEPLLN
TASEMDSKTMGPMVDYNNKDKINENUSNRYKCAVENYATLRLGLHAAAGGDPV
MRKVKLBSALAKREILRRHIDKVEVETIQAFTFYKVLIELYSGLLPPDVITIN
ADNFGMKRLSGPOEKSGRAGVYHLSYLSPEVSLMTSTTPALMVEERKAYD
TTADRWLANCGDLKGAETQISLFLMAYDIDSEFNADNATYPAWLAWFEERYD
LEDITCSHMLAFSRKREYMGVNNYWGGEKRTDTEFSFANVNEARRLTESRI
GKKTEDLLADESKKPAFYQLLYPVKGAELMNMHTIKGQFYROYVRAAANRLK
AOYKCHDSLLEITDGYNSLLDGKMHMSLKONYGTSYEMIPLMEERYVPVGP
LGIQABSENIDKSGMSPHLPAYSTYSRKSHWIDYNOGNGLSITPSEDMILISQ
KSGKTSANRITYISVDKYPAGEKRGQDIDISGQKENVUSVNPSPSPARTVOG
LYVEENGYISIPADPHRKESNDIMNSILPGIFGRSLQNGNPTAPLOMRAQV
RYEYPTFNAGYDYVTVLPTPLAERDYKLPEHTNSDITKYSVRIDDSISTST
SALEYSOIWDVYLKNCRVKSTLYVKKPGKHTLQRCDPGVVIQKIYIDLGKMS
YLGPOSTICN"
complement (3123..4292)
/locus_tag="BT3515"
complement (3123..4292)
/locus_tag="BT3515"
/codon_start=1
/transl_table=11
/product="putative glycosylhydrolase"
/protein_id="AA078621.1"
/db_xref="GI:29340831"
/trans_action="W1QYPMFLHCHYQMSRMBRPLILSLILIGVCOHKKAT
BKETVAGNTYTPLEERGAEPWALFHEGKYITQSGENKVIIMETDDITDLSHATQ
DWIPIPBNSYHMLABEIRINNKKWYIPAADGMNDNQLIVVENEANPEGVFV
MKGIQTDKDNMNAHASTPEHDQRYIMWGQKRIIDETOCIIYIAMKPMWLS
DRILSKREYECOWNPDSKTAIYIHVEAPOYFESKNDKACIFASAGSMTPY
ACQGLPADAKANLNPASWKSPPVLPQODPENNYPGSIGSTPSPCKEYMYLH
YACQIPNAPASDSRPRLOKIDMDQGMVLGTPKEEBPMARPSGSPHKQN"
4402..6315
/locus_tag="BT3516"
4402..6315
/locus_tag="BT3516"
/codon_start=1
/transl_table=11
/product="arabinan endo-1,5-alpha-L-arabinosidase A
precursor"
/protein_id="AA078622.1"
/db_xref="GI:29340832"
/translation="MKLRHLFLFCVLCQSVSISAONKSKLPKTSGNPIPGWYADBE
IVGDEWITPYTAAVDQIFMDAFSSKDLVAMWTKPYLSKENISMLRALMAVAV
IHANDKYFFFGANDIONNMLGIGVAVADNPAPKALGKPLIDKIVNAGOPIDQ
FVKKDDGQYVNYGKMGHCNMVAPADLLSTVPBEGTITKEVTPONTVEGPMKX
NGKTYFMSBSGSGDPYSAVAIADSPPEPFVVKTILODITNATGAGHSVHVC
GSDWYIILYARRPLGETANSRATCIDRMFNDGKLEVRMTFEGVAGSLTPVRA
KTYANPVINSLDPTIIGDDEYVYLYATESIKNVIHRSRLVWYVYGLATFTAT
RPNPEKGRVADPINKIDKRYVWYMSWGTEGTCIGIATADKREGFPTLGLAF
RSNEINONCIDPFIYEDNGKLYFMGSGFIYGTLELDGLLKKAPAEQIAGTAY
EGTYIHRGDGYVFAISIGCCBGLKSTYTTVGRSKNLFPGYVDDKGRMSNHHEI
LIQCKKAPVGGHSELVTDKGNDMFHYAVSVANDEGRVLMMDRQMGKMPFYK
AVSLETPADPDR"
6571..7143
/locus_tag="BT3517"
6571..7143
/locus_tag="BT3517"
/codon_start=1
/transl_table=11
/product="RNA polymerase Ecf-type sigma factor"
/protein_id="AA078623.1"
/db_xref="GI:29340833"

gene
CDS

/translation="MASDHTLSLSLMKRFLEBGSASQIYNQVQLFRYGLAYT
SDRELVDCTHDVFKIYTRAKLTPFDNIIVAYLMALKNTLFLNALKTSDDSLDA
DKEDQSEERSSTPETIYINKEOEKHNMVHMSLSLTTRQREIIVYRYIKMSIDE
ISKITDNNYQSVNSIORALGRANLFKRE"
7257..8459
/locus_tag="BT3518"
7257..8459
/locus_tag="BT3518"
/note="Tfrf homolog"
/codon_start=1
/transl_table=11
/product="putative anti-sigma factor"
/protein_id="AA078624.1"
/db_xref="GI:29340834"

gene
CDS

/translation="MINSBLKHTDFSRFTPEFLONDFFISSVRYKPTBEIOEFWRE
KSTPSNIDVPAAREYLETJSTSGDLSLOEGLMADIQTNIKDKIKHNFLLI
GLTPAASVAILVGSFPLKNVQALADINTFAVQITELPATEETLILAEDEVVL
KEKTTTIDVVALKABENISKEKLEALVNVQIVPRKRVIVLPPDSQKAVNAGTV
IYTFEEDKREIYDDEIYIIVARDERPFYTRKMNRAVLQKKNVAYASEPR
SVLAQGVQVETIQTAKIALAPQMSSVEKENISQVDEQMTISVNGLYCPNSAD
LGIYLRKLSITYYGINVEFDSLSKIKSGKIDKLKNEFTVINGLTFVAPISAYDQY
KYRVVKK"
8596..12051
/locus_tag="BT3519"
8596..12051
/locus_tag="BT3519"
/note="SusC homolog"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein, probably
involved in nutrient binding"
/protein_id="AA078625.1"
/db_xref="GI:29340835"

gene
CDS

/translation="MKKYNLILGLSLFPTVTLTSGDYFDQVDDRLSLBEIFKTR
DQALKYLSNVYFLIDPENQROVHETSLYRIPGWTGSDTEMVSSGNKALINNS
IDATENTWVLRKMSWISGTHESAVFTTYVQAPLTAEBRQWKAALAIAYEVL
VRYGVPLVEDVYALDTPSNELQSRSTVRCDFIVSELKEAQNALBEDASSDKT
TGVGRIDKAIQAQFIILALTYRASNLFGEGCTYVADNMVPGDRLPESQPADATTKAD
NOKVYECOKFADYGNFQMLMYTDKSGSVAGPABEFNFYRARIITLFSMGN
NKEMIHYRDNAGTMOYDRMKNKSGNTNDYRGSLIGATQTEMVADYRANMAGMSVGT
YAADVTPIINASGYKDKDSSSDYSKADGLVYAPGATRMNYVNRBRFVADITFSN
VNPT"
12064..13992
/locus_tag="BT3520"
12064..13992
/locus_tag="BT3520"
/note="Susd homolog"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein, probably
involved in nutrient binding"
/protein_id="AA078626.1"
/db_xref="GI:29340836"

/translation="MKKYNLILGLSLFPTVTLTSGDYFDQVDDRLSLBEIFKTR
DQALKYLSNVYFLIDPENQROVHETSLYRIPGWTGSDTEMVSSGNKALINNS
IDATENTWVLRKMSWISGTHESAVFTTYVQAPLTAEBRQWKAALAIAYEVL
VRYGVPLVEDVYALDTPSNELQSRSTVRCDFIVSELKEAQNALBEDASSDKT
TGVGRIDKAIQAQFIILALTYRASNLFGEGCTYVADNMVPGDRLPESQPADATTKAD
NOKVYECOKFADYGNFQMLMYTDKSGSVAGPABEFNFYRARIITLFSMGN
NKEMIHYRDNAGTMOYDRMKNKSGNTNDYRGSLIGATQTEMVADYRANMAGMSVGT
YAADVTPIINASGYKDKDSSSDYSKADGLVYAPGATRMNYVNRBRFVADITFSN
VNPT"


```

SKWFGTEGDTYVDFYSGNGKQAGNDFSTGYIVKXMDSGDRNONTVCYLLRLT
NIYFDYIEALAYVNDPSHADITKXNLLPERAGIGGYCTATLPRATTTSEMLKIOEXK
RIELSEFNCRYFDRWNGLTNEFFNFKVHGNNVYDGNFFYKTELTISNRFDRYFPP
IPGSEIDIDKXLVQNEGS"
14018..15307
/locus_tag="BT3521"
14018..15307
/locus_tag="BT3521"
/codon_start=1
/trans_table=11
/product="alpha-1,6-mannanase"
/protein_id="AA078627.1"
/db_xref="GI:29340837"
/translation="MMKQYIFSAICLVSAFCILSSGNDKEARPYTPDYRIVEEYTN
DPMKAYAEFNEHLDDQKPIYKSTADKAAVDENGAALWCPTWDMANNAKAK
AEQPTCKEKKFKQ.CDDLPAQNKAHYANPEFDNNENTGMFTDDIKMTVTTLARAYE
LPGVEEYLSSEBSFRWYSGEKVGTGSADBEKLGSGKMWMQPIKNPNENAD
HGKACINFTPTVAATLTVNNVPTGRTSDSHSYQTKQYLAKEQYADNAVMAVENV
DVTGGVADSRHNGNPAWKDHVQNSYIGASVLYKATGEKQYLDNAVMAVENV
TISGTPDLPEFTGAEQIYTAFAQYIAMLVYDCDQTOYIPVKNNINVMANREDKT
RDICGGDYTKLOVEGDAVESYSASGIPALMLLPPTDK"
15321..16493
/locus_tag="BT3522"
15321..16493
CDS
gene
Query Match 1.3%; Score 23; DB 1; Length 303446;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 411 ATTTCGCTTTCGAAGCAAC 433
Db 178526 ATTTCGCTTTCGAAGCAAC 178504

RESULT 8
VFA277287/c 2282 bp DNA linear PLN 07-DEC-2000
LOCUS
DEFINITION
Vicia faba partial enod18 gene, exon 1 and 5' UTR.
ACCESSION
A277287.1 GI:11602752
VERSION
A277287.1 GI:11602752
KEYWORDS
enod18 gene.
SOURCE
Vicia faba (fava bean)
ORGANISM
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
Vicia.
1
REFERENCE
1 Hohnjec,N., Kuster,H., Albus,U., Frosch,S.C., Becker,J.D.,
Puhler,A., Perlick,A.M. and Fruhling,M.,
The broad bean nodulin VENO18 is a member of a novel family of
plant proteins with homologies to the bacterial Mj0577 superfamily
Mol. Gen. Genet. 264 (3), 241-250 (2000)
JOURNAL
MEDLINE
20535713
PUBMED
11085263
2 (bases 1 to 2282)
REFERENCE
1 Kuester,H.
Direct Submission
Submitted (05-APR-2000) Kuester H., Lehrstuhl fuer Genetik,
Universitaet Bielefeld, Postfach 100131, NRW, GERMANY
LOCATION/Qualifiers
1..2282
/organism="Vicia faba"
/mol_type="genomic DNA"
/strain="Kleine Thueringer"
/db_xref="taxon:3906"
/tissue_lib="nucle cDNA library"
1948..2213
/gene="enod18"
1948..1958
/gene="enod18"
1957..2213
/gene="enod18"
FEATURES
source
gene
TATA_signal
mRNA

```

```

1997..2213
/gene="enod18"
/number=1
2039..2213
/gene="enod18"
/codon_start=1
/product="ENOD18 protein"
/protein_id="CAC18559.1"
/db_xref="GI:11602753"
/translation="WVEDRKVGVGIDFESKSNALKMAIVNADKDFEYLHINSNS
SDERSKLFKAKGS"
BASE COUNT 772 a 376 c 329 g 805 t
ORIGIN
Query Match 1.2%; Score 22; DB 8; Length 2282;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 887 TTAATATATGTTGTAC 908
Db 1271 TTAATATATGTTGTAC 1250

RESULT 9
AC133887 67271 bp DNA linear HTG 19-SEP-2002
LOCUS
DEFINITION
Homo sapiens clone CTD-2584F10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION
AC133887.1 GI:23196488
VERSION
AC133887.1 GI:23196488
KEYWORDS
HTG; HTGS_PHASEO.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 67271)
REFERENCE
1 Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens, clone CTD-2584F10
Unpublished
2 (bases 1 to 67271)
REFERENCE
1 Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Bonnaville,L., Boucknight,B.,
Camarata,A., Chang,J., Chazaro,B., Choepel,Y., Collumore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karalas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Melitini,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIPR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center Project name: L28217
Center Clone name: 2584_F_10
* NOTE: This record contains 81 individual

```

* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1
716 715: contig of 715 bp in length
816 815: gap of 100 bp
1549: contig of 734 bp in length
1550 1649: gap of 100 bp
2386: contig of 737 bp in length
2387 2486: gap of 100 bp
2487 3215: contig of 729 bp in length
3216 3315: gap of 100 bp
3316 4036: contig of 721 bp in length
4037 4136: gap of 100 bp
4137 4875: contig of 739 bp in length
4876 4975: gap of 100 bp
4976 5715: contig of 740 bp in length
5716 5815: gap of 100 bp
5816 6540: contig of 725 bp in length
6541 6640: gap of 100 bp
6641 7375: contig of 735 bp in length
7376 7476: gap of 100 bp
7476 8215: contig of 740 bp in length
8216 8315: gap of 100 bp
8316 9053: contig of 738 bp in length
9054 9153: gap of 100 bp
9154 9889: contig of 736 bp in length
9890 9989: gap of 100 bp
9990 10729: contig of 740 bp in length
10730 10829: gap of 100 bp
10830 11558: contig of 729 bp in length
11559 11658: gap of 100 bp
11659 12390: contig of 732 bp in length
12391 12490: gap of 100 bp
12491 13235: contig of 745 bp in length
13236 13335: gap of 100 bp
13336 14066: contig of 731 bp in length
14067 14166: gap of 100 bp
14167 14894: contig of 728 bp in length
14895 14994: gap of 100 bp
14995 15718: contig of 724 bp in length
15719 15818: gap of 100 bp
15819 16553: contig of 735 bp in length
16554 16653: gap of 100 bp
16654 17387: contig of 734 bp in length
17388 17487: gap of 100 bp
17488 18229: contig of 742 bp in length
18230 18329: gap of 100 bp
18330 19062: contig of 733 bp in length
19063 19162: gap of 100 bp
19163 19892: contig of 730 bp in length
19893 19992: gap of 100 bp
19994 20739: contig of 747 bp in length
20740 20839: gap of 100 bp
20840 21579: contig of 740 bp in length
21580 21679: gap of 100 bp
21680 22426: contig of 747 bp in length
22427 22526: gap of 100 bp
22527 23265: contig of 739 bp in length
23266 23365: gap of 100 bp
23366 24100: contig of 735 bp in length
24101 24200: gap of 100 bp
24201 24944: contig of 744 bp in length
24945 25044: gap of 100 bp
25045 25776: contig of 732 bp in length
25777 25876: gap of 100 bp
25877 26619: contig of 743 bp in length

26620 26719: gap of 100 bp
26720 27445: contig of 726 bp in length
27446 27545: gap of 100 bp
27546 28290: contig of 745 bp in length
28291 28390: gap of 100 bp
28391 29130: contig of 740 bp in length
29131 29230: gap of 100 bp
29231 29955: contig of 726 bp in length
29956 30055: gap of 100 bp
30056 30806: contig of 750 bp in length
30807 30906: gap of 100 bp
30907 31648: contig of 742 bp in length
31649 31748: gap of 100 bp
31749 32468: contig of 720 bp in length
32469 32568: gap of 100 bp
32569 33303: contig of 735 bp in length
33304 33403: gap of 100 bp
33404 34118: contig of 715 bp in length
34119 34218: gap of 100 bp
34219 34954: contig of 736 bp in length
34955 35054: gap of 100 bp
35055 35796: contig of 742 bp in length
35797 35896: gap of 100 bp
35897 36640: contig of 744 bp in length
36641 36740: gap of 100 bp
36741 37287: contig of 547 bp in length
37288 37388: gap of 100 bp
37389 38128: contig of 741 bp in length
38129 38228: gap of 100 bp
38229 38962: contig of 734 bp in length
38963 39062: gap of 100 bp
39063 39802: contig of 740 bp in length
39803 39902: gap of 100 bp
39904 40644: contig of 742 bp in length
40645 40744: gap of 100 bp
40745 41459: contig of 725 bp in length
41459 41563: gap of 100 bp
41570 42301: contig of 732 bp in length
42302 42401: gap of 100 bp
42402 43145: contig of 744 bp in length
43146 43245: gap of 100 bp
43246 43977: contig of 732 bp in length
43978 44077: gap of 100 bp
44078 44178: contig of 738 bp in length
44179 44816: gap of 100 bp
44816 44916: contig of 735 bp in length
44916 45650: contig of 720 bp in length
45651 45750: gap of 100 bp
45751 46470: contig of 720 bp in length
46471 46570: gap of 100 bp
46570 47203: contig of 733 bp in length
47203 47403: gap of 100 bp
47404 48138: contig of 735 bp in length
48139 48238: gap of 100 bp
48239 48965: contig of 727 bp in length
48966 49065: gap of 100 bp
49066 49798: contig of 733 bp in length
49799 49898: gap of 100 bp
49899 50631: contig of 733 bp in length
50632 50731: gap of 100 bp
50732 51463: contig of 732 bp in length
51464 52304: contig of 741 bp in length
52305 52404: gap of 100 bp
52405 53143: contig of 739 bp in length
53144 53243: gap of 100 bp
53244 53975: contig of 733 bp in length
53976 54075: gap of 100 bp
54076 54811: contig of 736 bp in length
54812 54911: gap of 100 bp
54912 55647: contig of 736 bp in length
55648 55747: gap of 100 bp
55748 56490: contig of 743 bp in length
56491 56590: gap of 100 bp

* 56591 57342: contig of 752 bp in length
 * 57343 57442: gap of 100 bp
 * 57443 58192: contig of 750 bp in length
 * 58193 58292: gap of 100 bp
 * 58293 59017: contig of 725 bp in length
 * 59018 59117: gap of 100 bp

Query Match 1.2% Score 22; DB 2; Length 67271;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Ox 974 TAAATTATGATTAAGTTGAT 995
 |||
 Db 26933 TAAATTATGATTAAGTTGAT 26954

RESULT 10
 HSDJ863C7/c 108653 bp DNA linear PRI 08-FEB-2001
 LOCUS Human DNA sequence from clone RP5-863C7 on chromosome 20p12.3-13.
 DEFINITION Contains the CSNK2A1 gene for casein kinase 2 alpha 1 polypeptide
 (EC 2.7.1.37), the 5' end of the gene for a novel protein similar
 to Drosophila CG17883, ESTs, STSs, GSSs and a Cpg island, complete
 sequence.
 ACCESSION AL049761
 VERSION AL049761.11 GI:12583640
 KEYWORDS HTG; casein kinase; Cpg island; CSNK2A1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 108653)
 Blakey, S.
 Direct Submission
 Submitted (07-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequests@sanger.ac.uk
 On Jan 27, 2001 this sequence version replaced gi:5738437.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em; EMBL; SW; SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information
 on the WORMPEP database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 20, constructed by the Sanger Centre Chromosome 20
 Mapping Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr20
 IMPORTANT: This sequence is not the entire insert of clone
 RP5-863C7. It may be shorter because we sequence overlapping
 sections only once, except for a 100 base overlap.
 The true left end of clone RP5-863C7 is at 1 in this sequence. The
 true left end of clone RP5-867F11 is at 79233 in this sequence. The
 true right end of clone RP5-852M4 is at 39224 in this sequence.
 This sequence was finished as follows unless otherwise noted: all
 chemistry was either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. RP5-863C7 is from the
 library RPCT-5 constructed by the group of Pieter de Jong. For
 further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2.
 Location/Qualifiers
 1..108653
 /organism="Homo sapiens"

/mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="20"
 /map="p12.3-13"
 /clone="RP5-863C7"
 /clone_11b="RPCT-5"
 6..316
 /note="AluSp repeat: matches 1..312 of consensus"
 318..602
 /note="AluO repeat: matches 1..308 of consensus"
 1396..1694
 /note="AluO repeat: matches 1..239 of consensus"
 2730..2848
 /note="AluO repeat: matches 353..471 of consensus"
 3386..3533
 /note="AluO/FRAM repeat: matches 156..302 of consensus"
 4442..4742
 /note="AluO repeat: matches 1..304 of consensus"
 4786..4965
 /note="L1MB7 repeat: matches 5955..6160 of consensus"
 4797..5802
 /note="Cpg island"
 /evidence="not_experimental"
 5300..5401
 /note="51 copies 2 mer cc 60% conserved"
 complement(5417..5612)
 /gene="dJ852M4.2"
 complement(5417..5612)
 /gene="dJ852M4.2"
 /product="dJ863C7.2 (Novel protein similar to Drosophila
 CG17883)"
 /note="match: ESTs: Em:BE276163 Em:BE409340 Em:BE276157
 Em:BE388289 Em:BE391114 Em:BE391493"
 /evidence="not_experimental"
 complement(5417..5486)
 /gene="dJ852M4.2"
 /note="continues in Em:AL121747 as dJ852M4.2"
 /codon_start=1
 /evidence="not_experimental"
 /product="dJ863C7.2 (Novel protein similar to Drosophila
 CG17883)"
 /protein_id="CAC28622.1"
 /db_xref="gi:12718359"
 /translation="MALRSAGDGTSGHWDGAEKA"
 5762..5920
 /note="WIR repeat: matches 83..251 of consensus"
 6030..6116
 /note="WER91C repeat: matches 14..104 of consensus"
 6141..6375
 /note="L2 repeat: matches 160..402 of consensus"
 6378..6453
 /note="Alu repeat: matches 2..77 of consensus"
 6779..7055
 /note="AluO repeat: matches 6..301 of consensus"
 7184..7480
 /note="AluX repeat: matches 14..310 of consensus"
 7829..8090
 /note="L1MB2 repeat: matches 5769..6032 of consensus"
 8190..8467
 /note="WER58B repeat: matches 21..318 of consensus"
 8503..8792
 /note="AluSp repeat: matches 1..290 of consensus"
 8819..9123
 /note="AluX repeat: matches 1..308 of consensus"
 9140..9262
 /note="FRAM C repeat: matches 1..121 of consensus"
 9268..10101
 /note="L1MB7 repeat: matches 1285..2098 of consensus"
 10102..10400
 /note="Alu repeat: matches 1..303 of consensus"
 10401..10623
 /note="L1MB7 repeat: matches 1060..1285 of consensus"
 10760..10942

FEATURES
 source

```

repeat_region /note="AluIo repeat: matches 1. 163 of consensus"
10943. 11411
repeat_region /note="LTR2 repeat: matches 1. 509 of consensus"
11412. 14527
repeat_region /note="HERVK22 repeat: matches 2332. 6837 of consensus"
14526. 14727
repeat_region /note="HERVK22 repeat: matches 699. 905 of consensus"
14716. 15515
repeat_region /note="HERVK22 repeat: matches 1. 808 of consensus"
15515. 15983
misc_feature /note="LTR2 repeat: matches 1. 509 of consensus"
complement(15889. 16375)
repeat_region /note="match: GSS: Em:AQ475722"
15985. 16108
misc_feature /note="AluIo repeat: matches 179. 302 of consensus"
complement(15993. 16331)
misc_feature /note="match: GSS: Em:AQ22357"
16081. 16361
misc_feature /note="match: STS: Em:G37195"
17097. 17135
repeat_region /note="L2 repeat: matches 2155. 2194 of consensus"
17136. 17414
repeat_region /note="AluIo repeat: matches 3. 300 of consensus"
17415. 17615
repeat_region /note="L2 repeat: matches 1964. 2155 of consensus"
17693. 18019
repeat_region /note="HALL repeat: matches 1194. 1528 of consensus"
18258. 18558
repeat_region /note="AluX repeat: matches 1. 299 of consensus"
19341. 19638
repeat_region /note="AluX repeat: matches 2. 299 of consensus"
19649. 19951
repeat_region /note="AluIo repeat: matches 1. 305 of consensus"
19982. 20269
repeat_region /note="AluIo repeat: matches 1. 290 of consensus"
20290. 20519
repeat_region /note="AluIo repeat: matches 59. 286 of consensus"
20791. 20953
repeat_region /note="AluIo repeat: matches 1. 168 of consensus"
20964. 21003
repeat_region /note="L2 repeat: matches 2627. 2701 of consensus"
21179. 21251
repeat_region /note="L2 repeat: matches 2627. 2701 of consensus"
21728. 22028
repeat_region /note="AluIo repeat: matches 1. 304 of consensus"
22211. 22325
repeat_region /note="MER91B repeat: matches 12. 123 of consensus"
22863. 23076
repeat_region /note="MER46A repeat: matches 1. 235 of consensus"
23315. 23359
repeat_region /note="L2 repeat: matches 2627. 2701 of consensus"
24196. 24443
misc_feature /note="match: STS: Em:G27906"
24895. 25060
misc_feature /note="match: GSS: Em:A2244707"
25073. 25276
misc_feature /note="match: GSS: Em:A2244707"
complement(25221. 25784)
misc_feature /note="match: GSS: Em:A0508904"
complement(25517. 25794)
gene /note="match: GSS: Em:AQ568163"
complement(25777. 26902)
gene /note="CSNR2A1"
complement(join(25777. 27157, 29457. 29543, 30508. 30656,
31759. 31859, 32961. 32962, 35335. 35445, 38800. 38883,
40802. 40861, 42336. 42386, 42914. 43015, 48159. 48310,
51532. 51741, 86753. 86902))
/mRNA
/product="dJ863C7.1.1 (casein kinase 2, alpha 1
polypeptide (EC 2.7.1.37))"
/note="isoform 1
match: cDNAs: Em:X54962 Em:J02853 Em:M55265 Em:U51866

```

```

Query Match 1.2% Score 22; DB 9; Length 108653;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 229 AACCTGAATCCATCATTTT 250
Db 9244 AACTGAATCCATCATTTT 9223

RESULT 11
AC121962
LOCUS
Mus musculus chromosome UNK clone Rp24-25515, WORKING DRAFT
SEQUENCE, 4 unordered pieces.
AC121962
AC121962.2 GI:22476135
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ACTIVERFIN.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 153537)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 153537)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (21-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 153537)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (25-AUG-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Aug 25, 2002 this sequence version replaced gi:21040086.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu; wustl.edu
Project Information
Center project name: M_B0255105
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phred; version 0.99019
Consensus quality: 152181 bases at least Q40
Consensus quality: 152299 bases at least Q30
Consensus quality: 152299 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 155087; sum-of-contigs
Quality coverage: 13.31 in Q20 bases; agarose-fp
Quality coverage: 11.42 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 14531: contig of 14531 bp in length
* 14532 14631: gap of unknown length
* 14632 152537: contig of 137906 bp in length
* 152538 152638: gap of unknown length
* 152638 153354: contig of 717 bp in length
* 153355 153454: gap of unknown length

```



```

* 22484 25382: contig of 2899 bp in length
* 25383 25482: gap of 100 bp
* 25483 28079: contig of 2597 bp in length
* 28080 28179: gap of 100 bp
* 28180 32156: contig of 3977 bp in length
* 32157 32256: gap of 100 bp
* 32257 35644: contig of 3388 bp in length
* 35645 35744: gap of 100 bp
* 35745 39479: contig of 3735 bp in length
* 39480 39579: gap of 100 bp
* 39580 43753: contig of 4174 bp in length
* 43754 50636: contig of 6783 bp in length
* 50637 50736: gap of 100 bp
* 50737 57916: contig of 7180 bp in length
* 57917 58016: gap of 100 bp
* 58017 63882: contig of 5866 bp in length
* 63883 63982: gap of 100 bp
* 63983 71647: contig of 7665 bp in length
* 71648 71747: gap of 100 bp
* 71748 78978: contig of 7231 bp in length
* 78979 79078: gap of 100 bp
* 79079 84661: contig of 5583 bp in length
* 84662 84761: gap of 100 bp
* 84762 90578: contig of 5817 bp in length
* 90579 90678: gap of 100 bp
* 90679 98969: contig of 8291 bp in length
* 98970 99069: gap of 100 bp
* 99070 111241: contig of 12172 bp in length
* 111242 111341: gap of 100 bp
* 111342 121097: contig of 9756 bp in length
* 121098 121197: gap of 100 bp
* 121198 130885: contig of 9688 bp in length
* 130886 130985: gap of 100 bp
* 130986 143583: contig of 12598 bp in length
* 143584 143684: gap of 100 bp
* 143684 160336: contig of 16853 bp in length.

FEATURES
source
1. 160336
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="20"
   /map="20"
   /clone="RP11-147121"
   /clone_lib="RP11 Human Male BAC"
1. 1332
   /note="assembly_fragment"
1433. 3055
   /note="assembly_fragment"
3156. 4227
   /note="assembly_fragment"
4328. 6154
   /note="assembly_fragment"
6235. 8472
   /note="assembly_fragment"
8573. 11191
   /note="assembly_fragment"
11292. 14390
   /note="assembly_fragment"
14491. 17075
   /note="assembly_fragment"
17176. 19638
   /note="assembly_fragment"
19339. 22383
   /note="assembly_fragment"
22484. 25382
   /note="assembly_fragment"
25483. 28079
   /note="assembly_fragment"
28180. 32156
   /note="assembly_fragment"
32257. 35644
   /note="assembly_fragment"

```

```

misc_feature /note="assembly_fragment"
35745. 39479
   /note="assembly_fragment"
misc_feature 39580. 43753
   /note="assembly_fragment"
misc_feature 43854. 50636
   /note="assembly_fragment"
misc_feature 50737. 57916
   /note="assembly_fragment"
misc_feature 58017. 63882
   /note="assembly_fragment"

```

```

Query Match 1.2% Score 22; DB 2; Length 160336;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 AACTGGAATCCATCATTTT 250
Db 111004 AACTGGAATCCATCATTTT 111025

```

```

RESULT 13
AC025089 165870 bp DNA linear PRI 11-JUN-2002
LOCUS Homo sapiens chromosome 8, clone RP11-275B19, complete sequence.
DEFINITION AC025089
ACCESSION AC025089.9 GI:21389296
VERSION HTG.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 165870)
Birtren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-275B19
Unpublished
2 (bases 1 to 165870)
Birtren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,U., Barna,N., Bastien,V., Bedalov,F.,
Boguslavsky,I., Bouckhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Goyette,N., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,J., Karatas,A.,
Klein,J., Laroque,K., Lamazares,R., Lander,T., Lebecky,J.,
Levine,R., Liu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheters,R.,
Meldrum,T., Menus,L., Mihova,T., Miranda,C., Mianga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Teefaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Vriel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
JOURNAL Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCES 3 (bases 1 to 165870)
AUTHORS Birtren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,I.,
Bouckhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cooke,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Fitzhugh,W., Gage,D.,
Galagan,J., Gardina,S., Ginde,S., Gird,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Laroque,K.,
Lamazares,R., Lander,T., Lebecky,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,D., Marquis,N.,

```

TITLE
JOURNAL
COMMENT

Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Neldrum, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neil, D., Oliver, J., Peterson, R., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S.,
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Strauss, N., Sudramanah, A., Talamas, J., Testaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (11-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 330 Charles Street, Cambridge, MA 02141, USA
On Jun 11, 2002 this sequence version replaced gi:17998733.

All repeats were identified using RepeatMasker:

Shit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www.seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L7911

Center clone name: 275_B_19

FEATURES

source

Location/Qualifiers
1. 165870
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone_lib="RP11-275B19"
/clone_lib="RP11 Human Male BAC"
2. 506
/rpt_family="L1"
510. 714
/rpt_family="THB1C"
717. 790
/rpt_family="L1"
792. 2114
/rpt_family="MSTRA-int"
2115. 2520
/rpt_family="MSTRA"
2544. 2726
/rpt_family="L1MB8"
2767. 4003
/rpt_family="L1MB8"
complement(4004..4372)
/rpt_family="THB1B"
complement(4373..4838)
/rpt_family="THB1B-int"
complement(4836..5274)
/rpt_family="THB1B-int"
complement(5277..5640)
/rpt_family="THB1B"
5642. 6034
/rpt_family="L1MB8"
6035. 6285
/rpt_family="ALUy"
6286. 6413
/rpt_family="L1MB8"
complement(6436..6798)
/rpt_family="L1MB5"
8239. 8274
/rpt_family="AT_rich"
10207. 10241
/rpt_family="(TA)n"
10786. 10807
/rpt_family="AT_rich"
11181. 11206

/rpt_family="(TA)n"
complement(11656..11884)
/rpt_family="L1MB2"
complement(12975..13256)
/rpt_family="42"
complement(13897..13957)
/rpt_family="L1PA2"
complement(20587..21007)
/rpt_family="MSTA"
21161..21189
/rpt_family="(TATC)n"
21228..21403
/rpt_family="MIR"
complement(21668..21768)
/rpt_family="MSTA"
22088..22268
/rpt_family="MIR"
complement(22451..23151)
/rpt_family="L1MB2"
complement(23890..23958)
/rpt_family="MIR"
complement(23984..24330)
/rpt_family="t199er1"
complement(24331..24448)
/rpt_family="FLAN_A"
complement(24449..26551)
/rpt_family="t199er1"
complement(28565..29190)
/rpt_family="HERV16"
29241..29362
/rpt_family="AT_rich"
29335..29398
/rpt_family="(TCTA)n"
30674..30696
/rpt_family="AT_rich"
30747..30800
/rpt_family="(CA)n"
31017..31037
/rpt_family="AT_rich"
31061..31088
/rpt_family="AT_rich"
complement(31950..32252)
/rpt_family="ALUy"
32597..33487
/rpt_family="MER4D"
34612..34646
/rpt_family="AT_rich"
complement(34653..35005)
/rpt_family="MER4A"
complement(35006..35100)
/rpt_family="MER4A"
complement(35105..35306)
/rpt_family="L1MA4A"
35460..35491
/rpt_family="(TTAA)n"
complement(36326..36519)
/rpt_family="L1MA4"
37109..37114
/note="<30 qual SNGL region"
complement(37161..37376)
/rpt_family="MLTIL"
complement(37407..40118)
/rpt_family="Char163"
complement(40124..40415)
/rpt_family="MLTIL"
40528..41815
/rpt_family="L1MA4A"
41814..41848
/rpt_family="AT_rich"
41988..42299
/rpt_family="AlusG"
43159
/note="probably G, possibly C"

unsure 43194, .43195
/note="probably CC"
repeat_region 43830, .44208
/rpt_family="LRR33"

Query Match 1.2%; Score 22; DB 9; Length 165870;
Best Local Similarity 100.0%; Pred.No.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 974 TAAATTCATTAAGTTGAT 995
Db 141362 TAAATTCATTAAGTTGAT 141383

RESULT 14 AC009336 175667 bp DNA linear PRI 17-OCT-2000
LOCUS Homo sapiens chromosome 2, clone RP11-387A1, complete sequence.
DEFINITION AC009336
VERSION AC009336.13 GI:10835353
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 175667)
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 2, clone RP11-387A1
Unpublished
2 (bases 1 to 175667)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Deyre, E., Devon, K., Dewar, K., Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, M., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Jones, C., Kam, L., Karatas, A., Lehoczy, J., Lien, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychalek, J., Naylor, J., Niloff, M., O'Connor, T., Roberts, D., Roy, A., Severy, P., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Tornuella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (16-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 175667)
Authors: Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bede, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campobasso, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Dewar, K., Dodge, S., Domingo, M., Doyle, M., Fehersor, J., Ferreira, P., Fitzhugh, M., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McEwan, P., Morrow, J., Naylor, J., Neuman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pletre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Tirrell, A., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

4 (bases 1 to 175667)
Authors: Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campobasso, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P., Fitzhugh, M., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A., LaRocque, K., Lamas, R., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., Menga, V., McEwan, P., Morrow, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychalek, J., Naylor, J., Niloff, M., O'Connor, T., O'Neill, D., Oliver, T. M., O'Neil, J., Peterson, K., Pletre, N., Pisan, C., Pollara, V., Raymond, C., Riback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Stange, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE

JOURNAL

Submitted (17-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Oct 17, 2000 this sequence version replaced gi:6984494.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: http://www-seg.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: 1935

Center clone name: 387_A_1

FEATURES

----- Location/Qualifiers

1..175667
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-387A1"
/clone_lib="RPCT-11 Human Male BAC"
complement(11..243)
/rpt_family="L1MEC"
complement(250..561)
/rpt_family="AluSg"
complement(600..816)
/rpt_family="L1MEC"
1457..1638
/rpt_family="AluSx"
complement(2510..2632)
/rpt_family="L2"
complement(2638..3056)
/rpt_family="MLT12"
3197..3503
/rpt_family="LTR16A1"
3602..3722
/rpt_family="LRR2B"
3728..3875
/rpt_family="AluY"
3881..4171
/rpt_family="LRR2B"
4986..5275
/rpt_family="AluYb"
5278..5326
/rpt_family="CA)n"
5818..5944
/rpt_family="L2"
5946..5986

repeat_region /rpt_family="AT_rich" 6023. .6330
 repeat_region /rpt_family="LTR37A" 6527. .6845
 repeat_region /rpt_family="Alub" 7059. 7110
 repeat_region /rpt_family="T-rich" 8666. .8898
 repeat_region /rpt_family="AT_rich" 12825. .12945
 repeat_region /rpt_family="FLAM A" 13097. .13390
 repeat_region /rpt_family="Alub" 13738. .13773
 repeat_region /rpt_family="AT_rich" 14420. .14720
 repeat_region /rpt_family="AlusX" 14848. .14898
 repeat_region /rpt_family="AT_rich" 14901. .15236
 repeat_region /rpt_family="AlusX" 15431. .15657
 repeat_region /rpt_family="MIR" 17847. .17869
 repeat_region /rpt_family="AT_rich" 18152. .18447
 repeat_region /rpt_family="Alub" 19551. .19767
 repeat_region /rpt_family="MIR" 19815. .19851
 repeat_region /rpt_family="MIR" 20616. .20913
 repeat_region /rpt_family="AlusX" 21282. .21310
 repeat_region /rpt_family="AT_rich" 21790. .21811
 repeat_region /rpt_family="L2" 22182. .22186
 repeat_region /rpt_family="Alub" 22260. .22594
 repeat_region /rpt_family="L2" 22594. .22594
 repeat_region /rpt_family="MER44D" 22595. .22736
 repeat_region /rpt_family="L2" 22821. .23239
 repeat_region /rpt_family="ML1K" 23634. .23790
 repeat_region /rpt_family="L2" 24009. .24368
 repeat_region /rpt_family="MER46B" 24965. .25287
 repeat_region /rpt_family="L2" 25382. .25989
 repeat_region /rpt_family="L1MCC" 25990. .26293
 repeat_region /rpt_family="Alub" 26294. .26954
 repeat_region /rpt_family="L1MCC" 26955. .27310
 repeat_region /rpt_family="THB1B" 27311. .27439
 repeat_region /rpt_family="L1MCC" 27440. .27636
 repeat_region /rpt_family="AlusG/X" 27665. .27927
 repeat_region /rpt_family="AlusG/X" 27932. .28395
 repeat_region /rpt_family="L1MCC" 28396. .28708

Query Match 1.2%; Score 22; DB 9; Length 175667;
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1037 TGAATATATGTTTCTATGTTA 1058
 Db 8376 TGAATATATGTTTCTATGTTA 8355

RESULT 15
 AC100799
 LOCUS
 DEFINITION Homo sapiens chromosome 8, clone RP11-1082P24, complete sequence.
 AC100799
 VERSION AC100799.2 GI:19551148
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 JOURNAL 1 (bases 1 to 194149)
 TITLE Homo sapiens chromosome 8, clone RP11-1082P24
 REFERENCE 2 (bases 1 to 194149)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 1 (bases 1 to 194149)
 2 (bases 1 to 194149)

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 194149)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Borkhagter, B., Brown, A., Camarata, J., Campolano, A., Chang, J., Chazaro, B., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fazo, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Lander, E., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (20-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

TITLE
 JOURNAL
 REFERENCE
 AUTHORS

Wed Nov '26 09:11:03 2003

us-09-831-083-1.rge

Page 17

Search completed: November 25, 2003, 18:16:37
Job time : 4401 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 ; Search time 2616 Seconds
(without alignments)
16565.321 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaaactctgacatcttga.....tagagcgatcaagctgaacc 1783

Scoring table: IDENTITY NUC
Gapop 10.0 ; Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

| | |
|-----|--------------|
| 1: | em_estba:* |
| 2: | em_estbun:* |
| 3: | em_estin:* |
| 4: | em_estm:* |
| 5: | em_estov:* |
| 6: | em_estpl:* |
| 7: | em_estro:* |
| 8: | em_hic:* |
| 9: | gb_est1:* |
| 10: | gb_est2:* |
| 11: | gb_hic:* |
| 12: | gb_est3:* |
| 13: | gb_est4:* |
| 14: | gb_est5:* |
| 15: | em_estfun:* |
| 16: | em_estom:* |
| 17: | em_gss_hum:* |
| 18: | em_gss_iv:* |
| 19: | em_gss_pln:* |
| 20: | em_gss_vit:* |
| 21: | em_gss_fun:* |
| 22: | em_gss_mam:* |
| 23: | em_gss_mus:* |
| 24: | em_gss_pro:* |
| 25: | em_gss_rtd:* |
| 26: | em_gss_png:* |
| 27: | em_gss_vrl:* |
| 28: | gb_gss1:* |
| 29: | gb_gss2:* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 114 | 6.4 | 1101 | 29 | CNS00EO7 |
| 2 | 104.6 | 5.9 | 1101 | 29 | CNS00EOVL |
| 3 | 103.6 | 5.8 | 1200 | 13 | BX415878 |
| 4 | 102 | 5.7 | 1200 | 13 | BX436510 |

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 5 | 101 | 5.7 | 1101 | 29 | CNS00EO7 |
| 6 | 101 | 5.7 | 1200 | 13 | BX415878 |
| 7 | 99.4 | 5.6 | 928 | 29 | CNS00DKY |
| 8 | 99 | 5.6 | 1101 | 29 | CNS00EVL |
| 9 | 98.2 | 5.5 | 1201 | 9 | AL565455 |
| 10 | 98.2 | 5.5 | 1201 | 13 | BX458623 |
| 11 | 97 | 5.4 | 1201 | 9 | AL536104 |
| 12 | 97 | 5.4 | 1201 | 13 | BX443774 |
| 13 | 96.6 | 5.4 | 1101 | 29 | CNS0021T |
| 14 | 95.8 | 5.4 | 1201 | 9 | AL536104 |
| 15 | 95.8 | 5.4 | 1201 | 13 | BX35216 |
| 16 | 95.2 | 5.3 | 1201 | 13 | BX458623 |
| 17 | 95 | 5.3 | 1200 | 13 | BX437758 |
| 18 | 94.8 | 5.3 | 661 | 29 | CNS020VT |
| 19 | 94.8 | 5.3 | 1098 | 13 | BX377526 |
| 20 | 94 | 5.3 | 1201 | 13 | BX395109 |
| 21 | 94 | 5.3 | 1201 | 13 | BX461824 |
| 22 | 93.4 | 5.2 | 1200 | 13 | BX436510 |
| 23 | 92.8 | 5.2 | 1092 | 29 | CNS020K7 |
| 24 | 92.6 | 5.2 | 1277 | 29 | CC253231 |
| 25 | 92.4 | 5.2 | 1056 | 13 | BX415058 |
| 26 | 92.2 | 5.2 | 1201 | 13 | BX443774 |
| 27 | 91.8 | 5.1 | 1098 | 13 | BX377526 |
| 28 | 91.2 | 5.1 | 1101 | 29 | CNS0039G |
| 29 | 91.2 | 5.1 | 1201 | 13 | BX366070 |
| 30 | 90.8 | 5.1 | 1201 | 13 | CNS0039G |
| 31 | 90.8 | 5.1 | 1201 | 13 | BX462207 |
| 32 | 90.6 | 5.1 | 1101 | 29 | CNS0021T |
| 33 | 90.4 | 5.1 | 994 | 29 | CNS04NOU |
| 34 | 90.2 | 5.1 | 854 | 9 | AL522840 |
| 35 | 90 | 5.0 | 1101 | 29 | CNS003BD |
| 36 | 89 | 5.0 | 1094 | 29 | CNS012FZ |
| 37 | 88.6 | 5.0 | 1148 | 9 | AL527799 |
| 38 | 88.6 | 5.0 | 1165 | 13 | BX383869 |
| 39 | 88.6 | 5.0 | 1201 | 13 | BX381083 |
| 40 | 88.4 | 5.0 | 1201 | 13 | BX426629 |
| 41 | 88 | 4.9 | 1147 | 14 | CD387645 |
| 42 | 87.8 | 4.9 | 1201 | 13 | BX366070 |
| 43 | 87.6 | 4.9 | 836 | 29 | CNS01100 |
| 44 | 87.2 | 4.9 | 1056 | 13 | BX415058 |
| 45 | 86.8 | 4.9 | 1200 | 29 | CNS016CO |

ALIGNMENTS

RESULT 1
LOCUS CNS00EO7
DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC: BACR29P01 of RPCL-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069440
VERSION AL069440.1 GI:4949583
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACR29B01"

/clone_id="RPCI-98"

/note="end : TET3"

BASE COUNT 366 a 66 c 104 g 351 t 214 others

ORIGIN

Query Match 6.4%; Score 114; DB 29; Length 1101;

Best Local Similarity 39.3%; Pred. No. 4.5e-05;

Matches 254; Conservative 98; Mismatches 291; Indels 4; Gaps 2;

QY 812 AATCTACAAAGAGAGATTTTAACTTAGAGAGTATGGAGTTAAAGACACA 871
 DB 409 AAGAGGAAAAAACAATKAAARGGGAKAAATATTKGAAAGATATAAAAA 468
 QY 872 CATTAGGGGAGGTGTTAAATTATGTTGTTGACCCAGCTACCTTGTAGCTTA 931
 DB 469 AATATATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 528
 QY 932 TAAGAAATTTGATATCATCATTAATTAATTAATTAATTAATTAATTAATTA 991
 DB 529 AATTTAAATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTA 588
 QY 992 GTATATTTAGAG---TTGAGAAACCAATATGCTGCTGCTGTTTGTATATGTT 1248
 DB 589 TTTTATTTTATTTATTTAAAGGAAAAAATTAATTAATTAATTAATTAATTA 648
 QY 1049 TTTCTATGTTACTTTCTTCAAGCTATATAAAAACTTTGATGCTTAATTTGAT 1108
 DB 649 TTAATATTAATTTATTAATTTATTAATTTATTAATTTATTAATTTATTAAT 708
 QY 1109 AAAAAAATGTTAATGAATTCGAATGATATGATTTCAAGTCCAAAATCCATCA 1168
 DB 709 AAAAAATTTAATTAATTAATTAATTAATTTTATTTATTAATTAATTAATTAAT 768
 QY 1169 TAGAAATTTAGTACAAAACGTAACTCAAAAATATCTCTTATTTAAATTTTCA 1228
 DB 769 AAATATTTTAAATTAATTAATTAATTTTATTTATTAATTAATTTATTAATTA 828
 QY 1229 ATATATAA-ATATCTGCTATTTTAAATTTTCAATTAATTAATTTGATGCTCT 1287
 DB 829 TTAATTTTATTAATTTTATTTTAAATTTTATTAATTTTATTAATTTATTAAT 888
 QY 1288 AGAATATCAACCAATATTAATTAATTTTATTTATTAATTTTGAATCTCT 1347
 DB 889 AAATTAATAAATAATTAATTTTAAATTTTATTTTAAATTTTGAATATAAAAA 948
 QY 1348 CAATATATCTGATATTTTATTAATTTTATTTGCTGATTTTCTATTTGTTAG 1407
 DB 949 AATATTAATTAATTAATTTTATTTTAAATTTTATTTTAAATTTTGAATATA 1008
 QY 1408 CTTATATCTTGGTCAAACTAGTAAATTCATATATATGAGTTGTGAAGG 1454
 DB 1009 AAAAATATTTGATTTTAAATTTTATTTTAAATTTTATTTTAAATTTTGAAT 1055

RESULT 2

CNS00EVL

LOCUS 1101 bp DNA linear GSS 04-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC:
 BACR29B23 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION

AL0659706.1 GI:4949849

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

Bf 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a

collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila

melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila

melanogaster BAC library was prepared by Kazutoyo Osogawa and

Aaron Mammotser in Peter de Jong's laboratory in the department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

NY. The library is named RPCI-98 and was constructed by partial

EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's

p1 and EST libraries. A more detailed description of the library

and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be

found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

/clone="BACR29B23"

/clone_id="RPCI-98"

/note="end : T7"

BASE COUNT 419 a 91 c 60 g 299 t 232 others

ORIGIN

Query Match 5.9%; Score 104.6; DB 29; Length 1101;

Best Local Similarity 34.8%; Pred. No. 0.00066;

Matches 218; Conservative 120; Mismatches 285; Indels 4; Gaps 1;

QY 830 ATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 889
 DB 459 AHTTATTTTAAATTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 518
 QY 890 AAATTAATGTTGTAACCACTACCTTATGATATTAATTAATTAATTTATCAT 949
 DB 519 AATTTTAAATTTTATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 578
 QY 950 CACATTAATTAATTTGCTTATTTAAATTAATTAATTAATTTGATCATTAAGATTG 1009
 DB 579 AATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 638
 QY 1010 AAACCAATATGCTGCTGCTGATTTTGAATTTTATTTTCAATTTTCTTCTCA 1069
 DB 639 AATTTAATTTTATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 698
 QY 1070 GCTTATTAATTAATTTGTAATGCTTAATTTGCTGGAATTAATTTGTAATG 1129
 DB 699 AAAAAAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 758
 QY 1130 AATGAATTAATGATTTTCAAGTCCAAAATCCATCAATTAATTTAGTCAAAAG 1189
 DB 759 AATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 818
 QY 1190 AATCAAAATATTTCTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTCTT 1245
 DB 819 AAAAAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAA 878

| | | | | |
|------------|--|---|---|------------------------|
| OY | | 1246 | TTTAAATTTAGACATAAATAATATTTCACCTGTGCAACTTTAGAAAATCCACAACATA | 1305 |
| Dd | | 879 | ATAAAATTTWTTTWTTTTWWAAWAATATAAAMAATAAAMAAAAAATAAATAAAMMW | 938 |
| OY | | 1306 | TTAATACTAGATAATTTTATTCTTAAPATATTGGATCTCTCAATATATCTGAATTTA | 1365 |
| Dd | | 939 | TWATATTTWTTATTTAAATTTWTTAAATTTWTTAAATTTWTTAAATTTWTTAAATTTWTTA | 998 |
| OY | | 1366 | TTTTATATTTTGTCATATTTTCTTAGTTTGTAGAGTTAACCTTAATCTTGTCACAC | 1425 |
| Dd | | 999 | WTAAATATATTTATTAAMWTATATTTTMAAAAWTAAATATATTAATAAAMWTAAATANA | 1058 |
| OY | | 1426 | TAGTAATTCATATATATAGTTTGAA | 1452 |
| Dd | | 1059 | AATATMTATATATATWTAATATMAAA | 1085 |
| RESULT 3 | | | | |
| LOCUS | | BX415878 | 1200 bp | mRNA |
| DEFINITION | | BX415878 Homo sapiens THYMUS Homo sapiens CDNA clone CSOCAP008YI04 | | linear EST 15-MAY-2003 |
| ACCESSION | | BX415878 | | |
| VERSION | | BX415878.1 GI:30765550 | | |
| KEYWORDS | | EST. | | |
| SOURCE | | Homo sapiens (human) | | |
| ORGANISM | | Homo sapiens | | |
| REFERENCE | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| AUTHORS | | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | | |
| TITLE | | 1 (bases 1 to 1200) | | |
| JOURNAL | | Li, W.B., Gruber, C., Jesse, J. and Polayes, D. | | |
| COMMENT | | Full-length cDNA libraries and normalization unpublished | | |
| | | Contact: Genoscope | | |
| | | Genoscope - Centre National de Sequencage | | |
| | | Bp 191 J.006 EVRY cedex - France | | |
| | | Email: seqreg@genoscope.cns.fr, Web : www.genoscope.cns.fr | | |
| | | Library was constructed by life technologies, a division of | | |
| | | Invitrogen. Contact : Feng Liang Email : fliang@life tech.com URL : | | |
| | | http://fulllength.invitrogen.com/Invitrogen Corporation 1600 | | |
| | | Paradey Avenue Genoscope sequence ID : CSOCAP008B50ZP1. | | |
| FEATURES | | | | |
| source | | Location/Qualifiers | | |
| | | 1..1200 | | |
| | | /organism="Homo sapiens" | | |
| | | /mol_type="mRNA" | | |
| | | /db_xref="taxon:9606" | | |
| | | /clone="CSOCAP008YI04" | | |
| | | /tissue_type="THYMUS" | | |
| | | /clone_lib="Homo sapiens THYMUS" | | |
| | | /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed | | |
| | | with a NotI-oligo(dT) primer. Five prime end enriched, | | |
| | | double-strand cDNA was digested with Not I and cloned | | |
| | | into the Not I and EcoRV sites of the pCMVSPORT 6 vector. | | |
| | | Library was not normalized." | | |
| BASE COUNT | | 428 a 70 c 85 g 44 t 170 others | | |
| ORIGIN | | | | |
| | | Query Match 5.8%; Score 103.6; DB 13; Length 1200; | | |
| | | Best Local Similarity 38.1%; Pred.No.0.00085; | | |
| | | Matches 271; Conservative 107; Mismatches 329; Indels 5; Gaps 3; | | |
| OY | | 885 | TGTTAAATTTATGTTGTATAACCACCACTACCTTTAGTAAGTATTAAGAATAATTTGA | 944 |
| Dd | | 466 | TATTAAATTTATTTATTTTAAAAATTTTAAATTTTAAATTTTAAATTTWTAATTTAA | 525 |
| OY | | 945 | ATCATACATTAATTAATTTGTCTTATTTAAATTAATGATTAAGTCTATCATTAAGAT | 1004 |
| Dd | | 526 | TTTTTAAATTTAAATTAATTAATTTTAAAAATTTTAAAAAATAAATAAATWTTTWT | 585 |
| OY | | 1005 | TGAGAAACCAAATAGTCTCGTCTGATTTTGAATTAATGTTTTCATAGTACTTTC | 1066 |
| Dd | | 586 | TTTAAATTTTATTTTATTTTWTWTWTATTTTATTAATTAATTTTAAATTAATTA | 645 |

| | | | | |
|------------|----|------|---|------|
| | QY | 1065 | TTCACGCCCTATATAAAAGCTTGTGAATGCATAAATTGATCTGGAAAAAATGCGTAATG | 1124 |
| | Dd | 646 | TATATAAAT- TAAWTRBAWMAAWMTAATTTTTTTTATATAGATTAATATATATATTT | 704 |
| | QY | 1125 | AATTCATACGAATATATGTAATTCCAAAGTCGCAAAATCCATCATATAGAATTTAGTACA | 1184 |
| | Dd | 705 | TTATTTTATTTAAATTTWWAMTTTTTTTWAAPAGAAAMTATTAATACATATAATATA- -TATA | 762 |
| | QY | 1185 | ACGCTAACGCAAAAATATTCCTATTTTATTTAAATTTTCAACAATATATAAATATCTCTT | 1244 |
| | Dd | 763 | WDCAATATAATATATWTATTAATATATTTTBMATAATWCAAAAATWAAAAAATTTTAAA | 822 |
| | QY | 1245 | ATTTTAAATTTTACATATATATAATTTATGACCTGTCACCTTGAGAAATACCACAACAT | 1304 |
| | Dd | 823 | AMTTTAAATTTWAAAAAAMAAATTTTAAAAAAATTTATTTTATTTTATTTTAAATTTAA | 882 |
| | QY | 1305 | ATTATACTAGATATTTTATTCCTAATATTTTGAGATCCTGCAATATATCGATATTT | 1364 |
| | Dd | 883 | AMTAAAAATTSAAAAAT--TTTTTWTTTATTTATTTTAAATTTTAAATTTTTTTTWWMA | 940 |
| | QY | 1365 | ATTATTAATTTGTGCATATTTCTTATATGTTTAGAGTTAACCTTATATCTTGTCAAA | 1424 |
| | Dd | 941 | ATATAAATMAAAMAAAAAATTTWWMAAMWTATATAAATATMTATMTAAATWMAAAMTWMA | 1000 |
| | QY | 1425 | CTAGTAATTCAATATATAGCTTTGTGAAGCACACATTGACACTCTTGAAACCTGGTTTA | 1488 |
| | Dd | 1001 | TTWAMWMTWMAAAATMTATTTTMMWTWMAAAAAAAATTTTMMWTWMAAMTTTTTMMATA | 1060 |
| | QY | 1485 | ACCTGTTGGATGTTTAAAGTATATAACATTCAGATATGACCAATCTTATATATAC | 1544 |
| | Dd | 1061 | TATTTATTAATAATTTTWMAMAMATTTATTAATTTTATTTATTAATAATAATTTAAATTTWT | 1120 |
| | QY | 1545 | TTCCCTTGTCTTTTAAAAAAGTGTCATGGAAGATGCTCATGTGTAAGTACA | 1596 |
| | Dd | 1121 | TTTMTTATTTAAAAATATAAATTTAMATTTTMMWATTTAAATWTATMTTTTMTA | 1172 |
| RESULT 4 | | | | |
| EX436510 | | | 1200 bp mRNA linear EST 15-MAY-2003 | |
| LOCUS | | | EX436510 Homo sapiens THYMUS.Homo sapiens cDNA CSOCAP002YK09 | |
| DEFINITION | | | 5-PRIME, mRNA sequence. | |
| ACCESSION | | | EX436510 | |
| VERSION | | | EX436510.1 GI:30770190 | |
| KEYWORDS | | | EST. | |
| SOURCE | | | Homo sapiens (human) | |
| ORGANISM | | | Homo sapiens | |
| REFERENCE | | | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | |
| AUTHORS | | | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. | |
| TITLE | | | 1 (bases 1 to 1200) | |
| JOURNAL | | | L.J.W.B., Gruber,C., Jesse,J. and Polayes,D. | |
| COMMENT | | | Full-length cDNA libraries and normalization Unpublished Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr Library was constructed by life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 3593.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSOCAP002AF05QP&cluster=5393.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Paradey Avenue Genoscope sequence ID : CSOCAP002AF05QP1. Location/Qualifiers 1..1200 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSOCAP002YK09" /feature_type="THYMUS" /clone_id="Homo sapiens THYMUS" | |
| FEATURES | | | | |
| SOURCE | | | | |

/note="vector: pcnvsport 6: 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcnvsport 6 vector. Library was not normalized."

BASE COUNT 422 a 116 c 144 g 411 t 107 others

ORIGIN

Query Match 5.7%; Score 102; DB 13; Length 1200;
Best Local Similarity 41.5%; Pred. No. 0.0013;
Matches 217; Conservative 56; Mismatches 250; Indels 0; Gaps 0;

810 AAAATACACAAAGAGAGATTTTAACTAGAGAGATGAGAGATTAAGACCA 869
576 AA 635
870 CACATTAAAGGGAGGTGTTAAATTAATGTTGTAACCACTACCTTAGTAGT 929
636 ATAAAAATTTTAAAAATTTAAAAATTTTAAAAATTTTAAAAATTTT 695
930 TATAGAAATTTGATGATCATCATTAATTAATTTGCTTATTAATTAATGATAG 989
696 TTTATATATATTTTATATATATATATATATATATATATATATATATAT 755
990 TTGATCATTAAGATGAGAAACCAATAGCTCTGCTGATTTTGAATTAATGTT 1049
756 ATTATATATATTTTAAAAATTTAAAAATTTTAAAAATTTTAAAAATTT 815
1050 TCTATGTTACTTTCTTAAAGCTATATATATATATATATATATATATAT 1109
816 TTTTATTTTATTTTAAAAATTTTAAAAATTTTAAAAATTTTAAAAATTT 875
1110 AAAAAATGTAAGATTCATAGAAATTTGATTTCAAGTCCAAATCCATCAT 1169
876 AAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 935
1170 AGAATTTAGTACAAAGCTACCAAAATTTCTTATTTAAATTTTACACATA 1229
936 TTTAAAAATTTATATATTTTAAAAATTTTAAAAATTTTAAAAATTTT 995
1230 TAAAAATTTCTTATTTTAAATTTTACATTAATTAATTTACCTGACCTTAG 1289
996 TTTATTTTAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 1055
QY 1290 AATACCAACCAATTAATTAATTTAGATTTTATTTATTTATTTAT 1332
Db 1056 AAAAAATTTTATTTTAAAAATTTTAAAAATTTTAAAAATTTTAAAAAT 1098

RESULT 5
CNS00807 1101 bp DNA linear GSS 04-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TBT3 end of BAC:
BACR2BP01 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL069440.1 GI:4949583
VERSION
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kanuoyo Osoegasa and Aaron Mammeter in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR2BP01"
/clone_11b="Rpci-98"
/note="end : TBT3"

BASE COUNT 366 a 66 c 104 g 351 t 214 others

ORIGIN

Query Match 5.7%; Score 101; DB 29; Length 1101;
Best Local Similarity 38.0%; Pred. No. 0.0018;
Matches 206; Conservative 80; Mismatches 256; Indels 0; Gaps 0;

906 ACCACACACTACCTTAGATGATATATAGAAATTTGATATCATCATTAATTAATG 965
1019 AAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 960
966 TCTTATTTAAAAATTTATGATTAAGTTGATCATTTAAGTTGAAACCAATAGCTC 1025
959 TACATATATTTTATTTTATTTTATTTTAAAAATTTAAAAATTTTAAAAATTT 900
QY 1026 GCTTGATTTTGAATTAATTTGTTTCTATGTTCTTCTCAAGCTATATTAATTA 1085
Db 899 TTTTATTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 840
QY 1086 TGTAAATGCTAAATTTGATGCGGAAAAAATGCTGATGAATTCATTAATTA 1145
Db 839 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 780
QY 1146 TTCAAGTCCAAATTCATGATTAATTTAGTAAATTTAGTAAATTTAGTAA 1205
Db 779 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 720
QY 1206 CTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1265
Db 719 TTTAAATTTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 660
QY 1266 TTAATTTAGCTGACCTTTAGATTAATTTAGTAAATTTAGTAAATTTAGT 1325
Db 659 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 600
QY 1326 TCTTAAATTTTGAATCTGATATATTTGATTTTATTTTATTTTATTT 1385
Db 599 AAAAAATTTAAATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 540
QY 1386 TTTTATGTTTATGATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1445
Db 539 TTTATTTAAATTTGATTTTATTTTATTTTAAATTTTAAATTTTAAATTT 480
QY 1446 TT 1447
Db 479 WT 478

RESULT 6
BX415878/c 1200 bp mRNA linear EST 15-MAY-2003
LOCUS BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS00AP008Y104
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION BX415878

| | | | | | |
|------------|-------|------|------|-------|------------|
| BASE COUNT | 419 a | 91 c | 60 g | 299 t | 232 others |
| ORIGIN | : | : | : | : | : |

http://www.genoscope.cns.fr/cgl-bin/cluster.cgi?seq=CS0DFO05BH09NP1&cluster=9232.f. Contact: Feng Liang Email: fliang@lifetech.com URL:

QY 1012 ACCAATAGTCTGCTGATTTTGAATTAATGTTTCTAGTGTACTTTCTTCAAGC 1071
 DB 575 AA 634
 QY 1072 CTATATATAAAGCTTGTAAGCTTAATGTATGCTGAGAAAAAATGTAATGATCA 1131
 DB 635 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 694
 QY 1132 TAGAATATATGATTTTCAAGTCAAAATCATCAATAGAAATTTAGTCAAAAGCTAA 1191
 DB 695 AA 754
 QY 1192 CTCAAAATATCTCTTATTTTAAATTTTCAACAAATATAAATAATCTCTTATTTTAA 1251
 DB 755 AAAAAAAAAADTTTAT 814
 QY 1252 ATTTTACAT 1311
 DB 815 TTTTATATATAAATAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 874
 QY 1312 CTATGATATTTTATTTCTTAATATTTTGAATCTTCAATATATCTGATATTTATTTAT 1371
 DB 875 TTTTATATAAATAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 934
 QY 1372 ATTTGCTCATATTTTCTTATGTTTGAAGTCAACCTTATCTTGTGCAACTAGTAA 1431
 DB 935 AAAAAAAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 994
 QY 1432 TTCAATATATAGTTTGTGAAGACATGACATCTTGAACATTTGTTTAACTGTG 1491
 DB 995 ATATATATATAAATAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1054
 QY 1492 TGAATATTTAAAGTATATAAATCATGATATATGACATCTTATATATCTTCTTT 1551
 DB 1055 ATTTTATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1114
 QY 1552 GTCTTTTAAATAAGTGTGATGAATAATGCTCTATGTTAGTGAAGTGTGCTGGCCT 1611
 DB 1115 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1174
 QY 1612 GTGTATATCATTTTCCATT 1630
 DB 1175 AA 1193

RESULT 13
 CENS0021J/c 1101 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL061936
 VERSION AL061936.1 GI:4940214
 SOURCE GSS.
 ORGANISM Drosophila melanogaster (fruit fly)
 KEYWORDS Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Genoscope.
 TITLES Direct Substition
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazuhiro Oosagawa and
 Aaron Mammose in Pieter de Jong's laboratory in the Department of

FEATURES
 source
 1..1101
 location/Qualifiers
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone_1ib="RPCI-98"
 /note="end : TET3"
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

BASE COUNT 631 a 7 c 28 g 289 t 146 others
 ORIGIN

Query Match 5.4%; Score 96.6; DB 29; Length 1101;
 Best Local Similarity 42.6%; Pred. No. 0.0064;
 Matches 258; Conservative 37; Mismatches 310; Indels 0; Gaps 0;

QY 954 TATATATATATGCTCTTATTTTAAATTTGATTAAGTGTATCATTAAGATTGAGAAAC 1013
 DB 916 TTAACCTTAAAAACATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 857
 QY 1014 CAATATAGTCTGCTGCTGATTTTGAATTAATTTCTATGTTTCTTCAAGCT 1073
 DB 856 AAAAAAAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 797
 QY 1074 ATATATAAAGCTTGTAAGCTTAATTTGATGCTGAGAAAAAATGTAATGATCAATA 1133
 DB 796 AAAAAAAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 737
 QY 1134 GAATATATGATTTTCAAGTCAAAATCCATCAATATAAATTTAGTCAAAAGCTAACT 1193
 DB 736 AAAAAAAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 677
 QY 1194 CAAAAATATCTCTTATTTTAAATTTTCAACATATATAAATAATCTTATATTTAAAT 1253
 DB 676 AAAAAAAAAATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 617
 QY 1254 TTTACAT 1313
 DB 616 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 557
 QY 1314 TAGATATTTTATTTCTTAATAATTTTGAATCTCTCAATATATCTGATATTTTATAT 1373
 DB 556 AAAAAAAAAAAAAAAAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAA 497
 QY 1374 TTGCTCATTTTCTTATGTTTGAAGTGAACCTTATATCTTGTGCAAACTGTAAT 1433
 DB 496 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 437
 QY 1434 CAATATATAGTCTTGTGAAGACATGACATCTTGAACATGTTTAACTGTG 1493
 DB 436 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 377
 QY 1494 GATGTTAAAGGTATATAAATCATGATATATGACATCTTATATATCTTCTTTGT 1553
 DB 376 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 317
 QY 1554 CTCTT 1558
 DB 316 TTTT 312

RESULT 14
 AL536104/c 1201 bp mRNA linear EST 31-MAY-2003
 LOCUS AL536104 Homo sapiens FETYL BRAIN Homo sapiens cDNA clone
 DEFINITION CSDF022Yc18 5-PRIME, mRNA sequence.

```

FEATURES
source
Location/Qualifiers
1..1201
/organism="Homo sapiens"
/mol_type="mRNA"
/seq_xref="taxon:9606"
/clone="CSDF023YCI8"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dG) primer. Five prime end
enriched. double-strand cDNA was digested with NotI and
cloned into the NotI and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT
359 a 119 c 146 g 409 t 168 others
ORIGIN

```

[illegible]

| | |
|------------|--|
| RESULT 15 | |
| EX335216/c | |
| LOCUS | |
| DEFINITION | EX335216 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens CDNA |
| ACCESSION | EX335216 |
| VERSION | EX335216 |
| KEYWORDS | EST. |
| SOURCE | Homo sapiens (human) |
| ORGANISM | Homo sapiens |

REFERENCE
 Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1201)
AUTHORS
 Li, W.-B., Gruber, C., Jeessee, J., and Polayes, D.
TITLE
 Full-length cDNA libraries and normalization
JOURNAL
 Unpublished
COMMENT
 Contact: genomecope
 Genomecope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genomecope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 714.f For more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0D1012BD05Qp&cluster=714.f>. Contact : Feng Liang Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com/INV0D1012BD05Qp1>.
 Faraday Avenue Genomecope sequence ID: CS0D1012BD05Qp1.
FEATURES
 source
 1..1201
 Location/Qualifiers

| | | |
|---------------------------|--|---------------------------------|
| | /organism="Homo sapiens" | |
| | /mol_type="mRNA" | |
| | /db_xref="taxon:9606" | |
| | /clone="CSDBD1012XG10" | |
| | /tissue_type="PLACENTA | COT 25-NORMALIZED" |
| | /clone_id="Homo sapiens PLACENTA COT 25-NORMALIZED" | |
| | /note="1st strand cDNA was primed with a NotI-oligo(dT) | |
| | primer. Five prime end enriched, double-strand cDNA was | |
| | digested with Not I and cloned into the Not I and EcoR V | |
| | sites of the pCMVSPORT 6 vector. Library was normalized." | |
| BASE COUNT | 391 a | 115 c 137 g 453 t 105 others |
| ORIGIN | | |
| Query Match | 5.4% | Score 95.8; DB 13; Length 1201; |
| Best Local Similarity | 43.7%; Pred. No. 0.0076; | |
| Matches 255; Conservative | 46; Mismatches 271; Indels 12; Gaps 2; | |
| Dy | 1164 TAAAAATATATTCTTTAAAMWATTTATAAATATAAAMWMAAATAAAAAAAAAATW | 1105 |
| Dy | 887 TTTAAATTAATGTGTGTAACACCACCACCTACTTAGTAGTATTATAGAAMAATGGTAT | 946 |
| Dy | 947 CATGCATTATATTTATTCCTCATTTAAATTAATGATGAAGAAGTGATCATTAGATG | 1006 |
| Dy | 1104 TTATATAMTAAATADANNAATAATATTTTTATATATATATATANAATTTATATATATAAAAAW | 1045 |
| Dy | 1007 AGAAAAACAATAGCTCGCTTGATTTTGAATTAATGTTTCTANGTACTTTCTT | 1066 |
| Dy | 1044 WTAAATATATTTAAAMWATTAAMWATTAAMWAAAATAMWTTTAAATATATTTTNAATATA | 985 |
| Dy | 1067 CAAGCCCTATTAATAAC-----TTTGATATGCTAATATGATGCTGAAAAAATG | 1111 |
| Dy | 984 AATAAATATTAABAAAATATATATATATTTTAAAAAAAAMWMAAAATATATAMWATATA | 925 |
| Dy | 1118 TGTAATGATCAATDAGAAATTTATGTGATTTTCAAAGTCCAAAATCATCATAGAAATTT | 1177 |

This Page Blank (uspto)